

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 18:40:02 ; Search time 29.83 Seconds  
(without alignments)  
670.241 Million cell updates/sec

Title: US-09-828-217-1  
Perfect score: 889  
Sequence: 1 MASTSYDCRVPMEDGDKRC.....SSAAPQLLIVLLISALIQ 180

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: A\_Geneseq\_032802:\*  
2: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT:\*  
3: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT:\*  
4: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT:\*  
5: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT:\*  
6: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT:\*  
7: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT:\*  
8: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT:\*  
9: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT:\*  
10: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT:\*  
11: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT:\*  
12: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT:\*  
13: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT:\*  
14: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT:\*  
15: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT:\*  
16: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT:\*  
17: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT:\*  
18: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT:\*  
19: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT:\*  
20: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT:\*  
21: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT:\*  
22: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	889	100.0	180	AA1980.DAT
2	889	100.0	180	AA1981.DAT
3	889	100.0	180	AA1982.DAT
4	889	100.0	180	AA1983.DAT
5	889	100.0	180	AA1984.DAT
6	889	100.0	180	AA1985.DAT
7	889	100.0	180	AA1986.DAT
8	889	100.0	180	AA1987.DAT
9	889	100.0	180	AA1988.DAT
10	889	100.0	180	AA1989.DAT
11	889	100.0	180	AA1990.DAT

12	889	100.0	193	AA1991.DAT	Human colon cancer
13	882	99.2	180	AA1992.DAT	Protein bound by A
14	843.5	94.9	197	AA1993.DAT	Protein encoded by A
15	649.5	73.1	143	AA1994.DAT	HML1.24 antigenic p
16	649	73.0	132	AA1995.DAT	Soluble HML1.24 ant
17	649	73.0	161	AA1996.DAT	HML1.24 antigenic p
18	579	65.1	147	AA1997.DAT	HML1.24 antigenic p
19	574.5	64.6	126	AA1998.DAT	HML1.24 antigenic p
20	473.5	53.3	155	AA1999.DAT	Human HML1.24 ant
21	104.5	11.8	756	AA1999.DAT	Human OREF ORF2369
22	102.5	11.5	756	AA1999.DAT	Human OREF ORF2369
23	102.5	11.5	1484	AA1999.DAT	Polypeptide #2 for
24	99.5	11.2	270	AA1999.DAT	Canine ribosome re
25	99.5	11.2	270	AA1999.DAT	Low density lipop
26	99.5	11.2	270	AA1999.DAT	Bovine LDL recepto
27	99.5	11.2	273	AA1999.DAT	Bovine LOX-1 polyp
28	99	11.1	426	AA1999.DAT	Low density lipop
29	99	11.1	426	AA1999.DAT	Human OREF ORF1960
30	99	11.1	694	AA1999.DAT	Human protein sequ
31	95.5	10.7	962	AA1999.DAT	Human transport-as
32	95.5	10.7	977	AA1999.DAT	Human colon cancer
33	95.5	10.7	977	AA1999.DAT	Amino acid sequenc
34	95.5	10.7	1003	AA1999.DAT	Human protein SEQ
35	95.5	10.7	1003	AA1999.DAT	Human protein SEQ
36	95	10.7	885	AA1999.DAT	AMUL chromosome in
37	94.5	10.6	111	AA1999.DAT	Gene #29 human sec
38	93.5	10.5	140	AA1999.DAT	H. pylori GHPD 363
39	93.5	10.5	1390	AA1999.DAT	Drosophila melanog
40	93.5	10.5	1411	AA1999.DAT	Nucleolar/endosoma
41	93.5	10.5	1886	AA1999.DAT	Rattus norvegicus
42	93	10.5	944	AA1999.DAT	Novel human diagno
43	93	10.5	1972	AA1999.DAT	Smooth muscle myos
44	92.5	10.4	210	AA1999.DAT	Bacterial general
45	92.5	10.4	488	AA1999.DAT	Amino acid sequenc

## ALIGNMENTS

RESULT 1	
ID	AA872703 standard; Protein; 180 AA.
AA872703	
AC	AA872703;
DT	06-DEC-1995 (first entry)
DE	Human membrane polypeptide for enhancing pre-B cell growth.
XX	
XX	Rheumatoid arthritis; diagnosis; pre-B cell growth; enhancement.
OS	Homo sapiens.
PN	W095I0536-A.
XX	
PD	20-APR-1995.
XX	
PF	14-OCT-1994; 94WO-U01732.
XX	
PR	15-OCT-1993; 93UP-0281622.
XX	
PA	(HIRA/) HIRANO T.
XX	
PI	Hirano T, Kaisho T;
XX	
DR	WPI; 1995-161738/21.
DR	N-PSDB; AAQ89606.
XX	
PT	Recombinant membrane protein enhancing pre-B cell growth - and monoclonal antibody recognising it and useful in the diagnosis of rheumatoid arthritis
PS	Claim 1; Page 29-30; 40pp; Japanese.



CC (especially 5-100) mg/kg body weight. The humanised antibody has low  
 CC antigenicity and is therefore effective therapeutically in humans.  
 XX  
 SO Sequence 180 AA;

Query Match 100.0%; Score 889; DB 19; Length 180;  
 Best Local Similarity 100.0%; Pred. No. 6.2e-78;  
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTSYDYCRVPMEDGDKRCKLLGIGILVLLIIVLGVPPLIIFTIKANSEACRGDLRAV 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 mastsydycrvpmedgdkrcklllgigilvllilvlgvppliftikansacrgdlrav 60  
 QY 61 MECRVNTHLLGOELTEAOGFODVEAOAATCNHTYMAIASIDAKAGQKKVEELGEI 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 mecrvnthlllgqelteaogfodveaagatcnhtymaiaidaekagqkkveelgei 120  
 QY 121 TTLNHKLQDASAEVERLRENOVLSVRIDAKKYPSSODSSAAAPOLLIVLLGSLALQ 180  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 121 tllnhklqdasaeverlrrenovlsvridakkyppsgdssaaapqllivllglsallq 180

## RESULT 4

AAV33202  
 AAV33202 standard; Protein: 180 AA.

AC AAV33202;  
 DT 22-NOV-1999 (first entry)  
 XX  
 DE Human Hm1.24 antigenic protein.

XX Antigenic protein; Hm1.24; splice variant; promoter; antirheumatic;  
 KW antirheumatic; bone marrow; tumour cell; drug development; treatment;  
 KW myeloma; rheumatoid arthritis; human.  
 XX  
 OS Homo sapiens.

XX WO9943803-A1.

XX 02-SEP-1999.

XX 25-FEB-1999; 99WO-JP00884.

XX 25-FEB-1998; 98JP-0060617.

XX 24-MAR-1998; 98JP-0093863.

XX (CHUS ) CHUGAI SEIYAKU KK.

XX Ohtomo T, Tsuchiya M, Koshihara Y, Kosaka M;

XX WPI; 1999-550869/46.

XX N-PSDB; AAZ09726.

PT Genomic DNA encoding Hm1.24 antigen protein as well as splicing  
 PT variants; useful e.g. in development of drugs for treating myeloma and  
 PT rheumatoid arthritis

XX Example 1; Fig 1-2; 83pp; Japanese.

XX This invention describes a novel human antigenic protein, Hm1.24,  
 CC its encoding nucleic acid, splice variants and promoter region. The  
 CC products of the invention have antirheumatic and antirheumatic activity.

CC The DNA of the invention is isolated from bone marrow tumour cells,  
 CC which can be used to study the expression of Hm1.24 antigen, promoter  
 CC activity of its promoter region, and in development of drugs in treating  
 CC e.g. myeloma and rheumatoid arthritis. This sequence represents the  
 CC human Hm1.24 antigenic protein described in the invention.

XX Sequence 180 AA;

Query Match 100.0%; Score 889; DB 20; Length 180;  
 Best Local Similarity 100.0%; Pred. No. 6.2e-78;  
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTSYDYCRVPMEDGDKRCKLLGIGILVLLIIVLGVPPLIIFTIKANSEACRGDLRAV 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 mastsydycrvpmedgdkrcklllgigilvllilvlgvppliftikansacrgdlrav 60  
 QY 61 MECRVNTHLLGOELTEAOGFODVEAOAATCNHTYMAIASIDAKAGQKKVEELGEI 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 mecrvnthlllgqelteaogfodveaagatcnhtymaiaidaekagqkkveelgei 120  
 QY 121 TTLNHKLQDASAEVERLRENOVLSVRIDAKKYPSSODSSAAAPOLLIVLLGSLALQ 180  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 121 tllnhklqdasaeverlrrenovlsvridakkyppsgdssaaapqllivllglsallq 180

## RESULT 5

AAV32765  
 AAV32765 standard; Protein: 180 AA.

AC AAV32765;

DT 25-OCT-1999 (first entry)

XX Soluble Hm1.24 antigenic protein sequence.

XX Hm1.24; antigenic protein; antibody; immunoassay; diagnosis; cancer;  
 KW immune disorder; immunotherapy monitoring.

XX Homo sapiens.

XX WO9943703-A1.

XX 02-SEP-1999.

XX 25-FEB-1999; 99WO-JP00885.

XX 25-FEB-1998; 98JP-0060613.

XX (CHUS ) CHUGAI SEIYAKU KK.

XX Koshihara Y, Ozaki Y;

XX WPI; 1999-518836/43.

XX N-PSDB; AAZ10917.

PT Immunoassay of anti-Hm1.24 antibody or soluble Hm1.24 antigen,  
 PT useful for diagnosis of immune disorders and cancer

XX Example 3; Fig 14-15; 138pp; Japanese.

XX This sequence represents a human soluble Hm1.24 antigenic protein. The  
 CC invention relates to an immunochemical assay of anti-Hm1.24 antibody by  
 CC use of a soluble Hm1.24 antigenic protein, or an immunochemical assay of  
 CC the soluble antigen by use of the antibody. The immunoassay of the Hm1.24  
 CC antigen or antibody is useful for diagnosis of immune disorders and  
 CC cancer, for monitoring of anti-Hm1.24 antibody immunotherapy, and for  
 CC assay of the antibody or antigen for investigative purposes, in  
 CC biological samples such as blood, serum, urine, milk, synovial fluid or  
 CC microorganism culture media. The method is sensitive down to 500 pg/ml  
 CC antibody.

XX Sequence 180 AA;

Query Match 100.0%; Score 889; DB 20; Length 180;  
 Best Local Similarity 100.0%; Pred. No. 6.2e-78;  
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTSYDYCRVPMEDGDKRCKLLGIGILVLLIIVLGVPPLIIFTIKANSEACRGDLRAV 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 mastsydycrvpmedgdkrcklllgigilvllilvlgvppliftikansacrgdlrav 60

QY	61	MECWNTHTLLQOELTEAOKGFQDVEAQAATCNHTWTWALMASLDAEKAQOKKVEELEGEI	120
Db	61	meccrwthlllqgelteagqyfgdveaqaatcnhtwtwalmasldaekagqkkveelegei	120
QY	121	TTLNKHLKODASAEVERLRRENOVLVSRIDKKYYPPSSQSSSAAPQLITVLIGLSALIQ	180
Db	121	ttlnhklqdasaseverlrrenqvlsvriadkkypssqsssaapqllivllglsaliq	180
RESULT 6			
AAV02576			
ID AAV02576 standard; Protein; 180 AA.			
XX			
AAV02576;			
16-JUL-1999 (first entry)			
XX	DE	HMI.24 antigenic protein.	
XX	XX	Reconstituted human antibody; peptide antigen HMI.24; framework region;	
KW	KW	complementary determining region; CDR; anti-HMI.24 antibody; myeloma;	
KM	KM	humanised antibody.	
OS	OS	Homo sapiens.	
PN	PN	W09918212-A1.	
XX	PD	15-APR-1999.	
XX	PF	02-OCT-1998; 98WO-JP04469.	
PR	XX	03-OCT-1997; 97JP-0271726.	
XX	PA	(CHUS ) CHUGAI SEIYAKU KK.	
XX	PI	Tsuchiya M;	
XX	DR	WPI; 1999-277273/23.	
XX	DR	N-PSDB; AAX59485.	
PT	Reconstituted human antibody useful in the treatment of myeloma		
Disclosure; Page 131-133; 256pp; Japanese.			
CC	The specification describes a reconstituted human antibody recognizing		
CC	the peptide antigen HMI.24. This human antibody contains natural human		
CC	framework regions modified by amino acid substitutions to provide		
CC	homogeneity with a previously designed framework region (which may		
CC	arise from a human or non-human source); and complementary determining		
CC	regions (CDR) derived from a non-human anti-HMI.24 antibody. The		
CC	reconstituted antibody is useful in the treatment of diseases in which		
CC	the surface antigen HMI.24 is implicated such as myeloma. The present		
CC	sequence represents HMI.24 antigenic protein.		
XX	Sequence 180 AA;		
SQ	Sequence 180 AA;		
QY	Query Match	100.0%; Score 889; DB 20; Length 180;	
QY	Best Local Similarity	100.0%; Pred. No. 6.2e-78;	
QY	Matches 180; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Db	1	MASTSYDCRYVMEQGRCKLLIGIIVLITVILGVPLITFTIKANSEACRGLRAV	60
QY	1	mastysdcryvmedgdkrciklllqglivlllilvlygplllftikanseacrgllrav	60
Db	61	MECRNVTHTLLQOELTEAOKGFQDVEAQAATCNHTWTWALMASLDAEKAQOKKVEELEGEI	120
QY	61	mecrnvthlllqgelteagqyfgdveaqaatcnhtwtwalmasldaekagqkkveelegei	120
Db	121	TTLNKHLKODASAEVERLRRENOVLVSRIDKKYYPPSSQSSSAAPQLITVLIGLSALIQ	180
QY	121	ttlnhklqdasaseverlrrenqvlsvriadkkypssqsssaapqllivllglsaliq	180

RESULT	7
AAV05484	
ID	AAV05484 standard; Protein; 180 AA.
XX	
AC	
XX	
AAV05484;	
XX	
DT	
XX	
07-JUL-1999	(first entry)
XX	
Potentiator for antibody against lymphoid tumour.	
DE	
XX	
Antibody potentiator; lymphoid tumour; lymphoma; cytotoxic antibody;	
KW	
multiple myeloma; acute B-lymphocytoma; chronic B-lymphocytoma;	
KM	
pre-B lymphoma; Burkitt's lymphoma; acute T-lymphocytoma; therapy;	
KX	
chronic T-lymphocytoma; PNTL.	
XX	
Homo sapiens.	
OS	
WO9918997-A1.	
PN	
22-APR-1999.	
PD	
XX	
PE	
14-OCT-1998; 98WO-JP04645.	
PR	
05-AUG-1998; 98JP-0222024.	
XX	
14-OCT-1997; 97JP-0280759.	
XX	
PA	
(CHUS ) CHUGAI SEIYAKU KK.	
PI	
Koishihara Y, Kosaka M:	
XX	
WP1; 1999-277447/23.	
DR	
N-PSDB; AAX36561.	
XX	
Potentiation of antibody treatment of lymphoma with biological	
PT	
response modifier	
XX	
Claim 1; Page 37-38; 62pp; Japanese.	
PS	
This sequence represents a potentiator for an antibody against	
CC	
lymphoid tumour.	
CC	
The invention relates to a method for the treatment of lymphoma, in which	
CC	
a cytotoxic antibody is potentiated by administration of a biological	
CC	
response modifier. The method can be used for treatment of lymphomas and	
CC	
multiple myelomas which are resistant to conventional treatment, such as	
CC	
acute B-lymphocytoma, chronic B-lymphocytoma, pre-B lymphoma, Burkitt's	
CC	
lymphoma, acute T-lymphocytoma, chronic T-lymphocytoma, and PNTL.	
XX	
Sequence 180 AA;	
SO	
Query Match	100.0%; Score 889; DB 20; Length 180;
Best Local Similarity	100.0%; Pred. No. 6,2e-78;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	
1	MASTSVDCRYVMEGGDRCKLLIGILIVILLIYLGVPLLIIFTKANSEKRGGLRAY 60
Db	1 mastsvdycryvmeggdgdrcklllgilivillivlpglliflikanseacrdglrav 60
OY	61 MECRVNTHILODELTEAOKGFQDVEAOATCNHTYMALMASIDAEGAOKKVEELEGERI 120
Db	61 mecrvnthilqgelteaqdgfgdveagaatcnhtymalmasidaekagqkveelegerl 120
OY	121 TTLNHKLODASAEVERLRRENNOVLVSRIADKKYYRSSQDSSAAPQLILVILGISALIQ 180
Db	121 tflnhklqdaseverlrrenngvlsvriadkkypssqdsassaapqlilvliglsalig 180
RESULT	8
AAV07250	
ID	AAV07250 standard; Protein; 180 AA.
XX	

AC AAY07250;  
 XX  
 DT 06-JUL-1999 (first entry)  
 XX  
 DE BST-2 protein.  
 XX  
 KM Mouse; BST-2; monoclonal antibody; RS38; myeloma; cytotoxic activity.  
 XX  
 OS Mus sp.  
 XX  
 PN JP11092399-A.  
 XX  
 PD 06-APR-1999.  
 XX  
 PF 24-SEP-1997; 97JP-0274960.  
 XX  
 PR 24-SEP-1997; 97JP-0274960.  
 XX  
 PA (CHUS ) CHUGAI PHARM CO LTD.  
 XX  
 DR WPI: 1999-283503/24.  
 DR N-PSDB; AAX29996.  
 XX  
 PT An agent for treating myeloma -includes an antibody and has  
 cytoxic activity  
 AS  
 PS Claim 1; Page 10; 13pp; Japanese.  
 CC This sequence represents the mouse BST-2 protein which is used to raise  
 CC antibodies, especially the monoclonal antibody RS38. The antibody can be  
 CC used in compositions to treat myelomas when the antibody is associated  
 CC with a cytotoxic activity.  
 CC  
 SO Sequence 180 AA;

Query Match 100.0%; Score 889; DB 20; Length 180;  
 Best Local Similarity 100.0%; Pred. No. 6.2e-78;  
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTSYDYCRVPMEDGDKRCKLLGIGLVLLIIVIGVPLIIFTIKANSFACDGLRAV 60  
 |||  
 Db 1 mastsydycrvpmedgdkrcklllgilvllivllgvpiliftikansacrdglrav 60  
 QY 61 MECNNVTHLLOEELTEAOKGFQDVEAQAATCNHTVMALMASLDAEKAQGGKVEELGEI 120  
 |||  
 Db 61 mecnnvthllqgelteagkgfqedvqaatcnhtvmalmasldeakagqgkveelgei 120  
 QY 121 TTLNHHKLDASAEYERLRRENOVLSVRIADKKYPPSSODSSSAAPOLLIVLGLSALLQ 180  
 |||  
 Db 121 ttlhkhkldasaeeverlrrenovlsvriadkkypssodsssaapqllivlglisallq 180

RESULT 9  
 ID AAY53273  
 AAY53273  
 AC AAY53273;  
 DT 21-JUL-2000 (first entry)  
 XX  
 DE Human HML.24 antigen protein sequence SEQ ID NO:2.  
 XX  
 KM Human; HML.24 antigen protein; detection; plasmocytoma;  
 KM multiple myeloma; plasmocytic leukaemia; extramedullary plasmocytoma;  
 XX multiple myeloma; asymptomatic myeloma.  
 OS Homo sapiens.  
 XX  
 PN WO200017395-A1.  
 XX  
 PD 30-MAR-2000.  
 XX

PF 20-AUG-1999; 99WO-JP04502.  
 XX  
 PR 18-SEP-1998; 98JP-0264593.  
 XX  
 PA (CHUS ) CHUGAI SEIYAKU KK.  
 XX  
 PI Kawai S, Kolshibara Y, Kosaka M;  
 XX  
 DR WPI: 2000-283616/24.  
 DR N-PSDB; AAA13654.  
 XX  
 PT Detection or measurement of plasmocytomas, applicable for early  
 PT diagnosis of e.g. multiple myeloma and plasmocytic leukemia, using a  
 PT polynucleotide which is expressed specifically or strongly in  
 PT plasmocytomas  
 PS  
 PS Disclosure; Page 15-16; 20pp; Japanese.  
 XX  
 CC A method has been developed for detecting or measuring plasmocytomas in  
 CC a sample at an early stage of disease development. The method comprises  
 CC amplifying a polynucleotide which is expressed specifically or strongly  
 CC in plasmocytomas before quantifying the amplification product by  
 CC comparing with results obtained with a control sample. The method is for  
 CC detecting or measuring plasmocytomas, applicable for early diagnosis of  
 CC e.g. multiple myeloma, plasmocytic leukaemia, isolated plasmocytoma,  
 CC extramedullary plasmocytoma, multiple plasmocytoma obtained from smoking  
 CC or asymptomatic myeloma. The present sequence represents human HML.24  
 CC antigen protein, which is expressed in plasmocytomas and so can be used  
 CC in the method of the invention.  
 CC  
 SO Sequence 180 AA;

Query Match 100.0%; Score 889; DB 21; Length 180;  
 Best Local Similarity 100.0%; Pred. No. 6.2e-78;  
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTSYDYCRVPMEDGDKRCKLLGIGLVLLIIVIGVPLIIFTIKANSFACDGLRAV 60  
 |||  
 Db 1 mastsydycrvpmedgdkrcklllgilvllivllgvpiliftikansacrdglrav 60  
 QY 61 MECNNVTHLLOEELTEAOKGFQDVEAQAATCNHTVMALMASLDAEKAQGGKVEELGEI 120  
 |||  
 Db 61 mecnnvthllqgelteagkgfqedvqaatcnhtvmalmasldeakagqgkveelgei 120  
 QY 121 TTLNHHKLDASAEYERLRRENOVLSVRIADKKYPPSSODSSSAAPOLLIVLGLSALLQ 180  
 |||  
 Db 121 ttlhkhkldasaeeverlrrenovlsvriadkkypssodsssaapqllivlglisallq 180

RESULT 10  
 ID ABB50295  
 ABB50295  
 AC ABB50295;  
 DT 08-FEB-2002 (first entry)  
 XX  
 DE Bone marrow stromal antigen (BST-2) ovarian tumour marker protein, #80.  
 XX  
 KM Ovarian tumour marker gene; human; overexpression; upregulation;  
 KM epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;  
 KM identification; serous cystadenoma; borderline serous tumour;  
 KM serous cystadenocarcinoma; mucinous cystadenocarcinoma;  
 KM mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;  
 KM undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;  
 KM adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;  
 KM immune response pathway; cell proliferation regulation; protein folding;  
 KM membrane localised; secreted; therapeutic target; cytostatic;  
 KM gene therapy; vaccine.  
 XX  
 OS Homo sapiens.  
 XX



AC		AAG73947;			
XX					
DT	03-SEP-2001	(first entry)			
DE					
XX					
XX					
KW		Human colon cancer antigen protein SEQ ID NO:471.			
XX					
XX		Human: colon cancer; colon cancer antigen; diagnosis; detection; coloectal carcinoma; chromosome 19.			
OS	Homo sapiens.				
PN	WO200122920-A2.				
PD					
XX	05-APR-2001.				
PF					
XX	28-SEP-2000; 2000WO-US26524.				
PR					
PR	29-SEP-1999; 99US-0157137.				
PR	03-NOV-1999; 99US-0163280.				
PA					
XX	(HUMA-) HUMAN GENOME SCI INC.				
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;				
XX					
PT	WPI: 2001-235357/24.				
N-	N-PDB: AAH33378.				
XX					
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating coloectal cancers - Claim 11; Page 6512-6513; 9803pp; English.				
PS					
CC	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cyrostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate p expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patients own production of P. Additionally, N may be used to produce the colon cancer-associated ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of coloectal carcinomas and cancers. AAH37196 to AAH37204 and AAB7789 represent sequences used in the exemplification of the present invention.				
CC	N.B. Pages 66 to 662 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922.				
Sequence	193 AA:				
Query Match	100.0%; Score 889; DB 22; Length 193;				
Best Local Similarity	100.0%; Pred. No. 6.8e-78;				
Matches	180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
OY	1 MASTSYDRCVPMDGDKRKLLLGIGIVLIIIVILGYPLIIFTIKANSEACRDGLRAY 60				
Dd					
Db	14 mastsydyrcvpmedgdkrklllgigilvillivlgyplifftkanseacrdglray 73				
OY	61 MCGRNVTHILOELLEAOKGFODVEFAQATCTCHTWALMASIDAEKAOCKVYEELGEET 120				
Dd					
OY	74 mccrntvthllqelileaqdgigdveagaatcbtymalnaidaekagqqkvveeleget 133				
OY	121 TTLNRKLDDASAEVERLRRENQVLAVRIDKKRYFPSSODSSAAAAPOLLITVLLGSALLQ 180				
Dd					
OY	134 tllnhklqdaseverellrrengylsvriadkkrypsqsdaapqlilvllglsallq 193				
RESULT	13				

ID	AAW7292	standard; Protein; 180 AA.
XX	AAW7292;	
AC	AAW7292;	
XX		
DT	14-DEC-1998	(first entry)
XX		
DE	Protein bound by Anti-HM1.24 antibody.	
XX		
KW	Anti-HM1.24; antibody; lymphocyte activation inhibitor;	
KW	autoimmune disease; organ transplant; allergy.	
XX		
OS	Homo sapiens.	
XX		
PN	WO9837913-A1.	
XX		
PD	03-SEP-1998.	
XX		
PF	27-FEB-1998;	98WO-JP00831.
XX		
PR	28-FEB-1997;	97JP-0045663.
XX		
PA	(CHUS ) CHUGAI SEIYAKU KK.	
XX		
PI	Koishihara Y;	
XX		
DR	WPI; 1998-480937/41.	
DR	N-PSDB; AAV59114.	
XX		
PT	Lymphocyte activation inhibitor comprises antibodies, particularly	
PT	anti-HM1.24 antibody - for preventing and treating auto-immune	
XX	diseases, rejection reactions in organ transplant or allergy	
PS	Disclosure; Page 38-39; 53pp; Japanese.	
XX		
CC	The Anti-HM1.24 antibody can be used in the production of lymphocyte	
CC	activation inhibitors. These inhibitors can be used for the prevention	
CC	and treatment of autoimmune diseases, rejection reactions in organ	
CC	transplant or allergy. Administration is non-oral, e.g. by intra	
CC	venous and intramuscular injection, local or systemic.	
XX		
SQ	Sequence	180 AA;
Query Match	99.2%; Score 882; DB 19; Length 180;	
Best Local Similarity	99.4%; Pred. No. 3e-77;	
Matches 179; Conservative	0; Mismatches 1; Indels 0; Gaps	0;
OY	1 MASTSYDYCRVPEDGDKRCKLLGIGILLVLLIIVILGVPLITPTKANSEACRDLRAV	60
DB	1 mastsydycrvpmedgdkrccklllgllvlllilvllgvplllftkansaecrdglrav	60
OY	61 MCSRNVTHLLOELFEAOKGFODVEAOAATCMTVMALMASLDARKAOSGRKVEELEGEI	120
DB	61 mecrnvthllqelcleaagqldvyaqaactchvmaimasldaeeaqgqkvveelegei	120
OY	121 TTLNHRKLDASAEVRLRRENOVLVRIADKKRYTSSODSSAAAPOLILVILGHSALLO	180
DB	121 tllnhrklqdasaevrllrrenqvlsvriadkkrytssodssaaapqllilvllglsallq	180
RESULT 14		
AAW36951		
ID	AAW36951	standard; Protein; 197 AA.
XX		
AC	AAW36951;	
XX		
DT	12-MAY-1998	(first entry)
XX		
DE	Protein encoded by clone O238_1.	
XX		
KW	Human; secreted protein; molecular weight marker; genetic fingerprinting;	
KW	antibody production; nutritional supplement; therapy; clone O238_1;	

KM dendritic cell.  
 XX Homo sapiens.  
 OS  
 XX MO9740069-A2.  
 PN  
 XX 30-OCT-1997.  
 PD  
 XX 14-APR-1997: 97MO-0506134.  
 XX PF  
 XX 19-APR-1996: 96US-0635311.  
 XX PR  
 XX (GEM ) GENETICS INST INC.  
 PA  
 XX  
 PT Jacobs K, Lavallie ER, Mccoy JM, Merberg D, Racie LA;  
 DR Spaulding V;  
 DR WPI: 1997-535776/49.  
 DR N-PSDB: AAV00426.  
 XX  
 XX Isolated nucleic acid clones from ATCC 98028 encode novel secreted  
 PT proteins - having many potential uses, e.g. as immunomodulators,  
 PT cell proliferation or differentiation inhibitors or haematopoiesis  
 PT regulators  
 XX  
 XX Claim 36; Page 77; 114pp; English.  
 XX  
 CC This sequence is a protein encoded by clone 0238.1, which is a  
 CC polynucleotide of the invention. The DNA encoding this sequence was  
 CC isolated from a human dendritic cell cDNA library. The polynucleotide,  
 CC which encodes a secreted protein, can be used, e.g. as a tissue or  
 CC molecular weight marker, in genetic fingerprinting, to raise anti-protein  
 CC or anti-DNA antibodies and in interaction trap assays. The protein can be  
 CC used to assay biological activity, raise antibodies for use in  
 CC immunoassays, as a marker, to identify inhibitors of its interactions and  
 CC as a nutritional supplement. It may also have a very wide range of  
 CC therapeutic and biological activities (no examples are given to support  
 CC this), e.g. cytokine or modulator of cell proliferation and  
 CC differentiation, immunostimulant or immunosuppressant, haematopoiesis  
 CC regulator, bone, cartilage, tendon, ligament and/or nerve tissue growth  
 CC stimulator, follicle inhibitor/stimulator, chemotactic/chemokinetic,  
 CC haemostatic, thrombolytic or anti-inflammatory agent, antimicrobial,  
 CC biorrhythm, metabolism or behaviour modifier, anti-depressant or analgesic  
 CC or psoriasis treatment.  
 CC  
 SQ Sequence 197 AA:  
 30  
 Query Match 94.9%; Score 843.5; DB 18; Length 197;  
 Best Local Similarity 97.7%; Pred. No. 1.7e-73;  
 Matches 173; Conservative 1; Mismatches 2; Indels 1; Gaps 1;  
 QY 1 MASTSDYCVPMEDGKRCKLLGIGLIVLITVITIGVPLIFETJKANSEACRDGIRAV 60  
 DB 1 mastsdycvpmedgkrcklllgilvllitvllivllivpdliflkanseacdrlrav 60  
 QY 61 MECCRNTHLLQOELTEAOKGFQDVEAQAATCNHTVMAALASLDAERAKOGKVEELEGEL 120  
 DB 61 meccrnthlllqelteaokgfgdveaq-ahcnhtmalmaslaekagqkveelegeel 119  
 QY 121 TTTNHLKQDASAEVERLRRENOVLSVRIADKKYPPSSQSSAAAPQLLIVLIGLSA 177  
 DB 121 ttnhlkqdasaeverlrrenovlsvriadkkypssqssaaapqllivlignls 176  
 RESULT 15  
 ID AAY32763 standard; Protein; 143 AA.  
 XX  
 AC AAY32763;  
 XX  
 DT 25-OCT-1999 (first entry)  
 XX

DE HML.24 antigenic protein/HA peptide fusion protein sequence.  
 XX  
 XX HML.24; antigenic protein; antibody; immunoassay; diagnosis; cancer;  
 KM immune disorder; immunotherapy monitoring.  
 XX  
 XX Homo sapiens.  
 OS  
 XX Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FT Misc-difference 7  
 FT /label= encoded by GAC  
 XX  
 XX MO9943703-A1.  
 XX  
 XX 02-SEP-1999.  
 XX PD  
 XX 25-FEB-1999: 99MO-JP00885.  
 XX PF  
 XX 25-FEB-1998: 98JP-0060613.  
 XX PR  
 XX (CHUS ) CHUGAI SEIYAKU KK.  
 XX PA  
 XX Koishihara Y, Ozaki Y;  
 XX PI  
 XX WPI: 1999-518836/43.  
 XX DR  
 XX N-PSDB: AA210915.  
 XX DR  
 XX  
 PT Immunoassay of anti-HML.24 antibody or soluble HML.24 antigen,  
 PT useful for diagnosis of immune disorders and cancer  
 PT  
 XX  
 XX Disclosure; Page 77-78; 138pp; Japanese.  
 XX  
 CC This sequence represents a fusion protein of the human soluble HML.24  
 CC antigenic protein and a HA peptide. The invention relates to an  
 CC immunochemical assay of anti-HML.24 antibody by use of a soluble HML.24  
 CC antigenic protein, or an immunochemical assay of the soluble HML.24  
 CC use of the antibody. The immunoassay of the HML.24 antigen or antibody is  
 CC useful for diagnosis of immune disorders and cancer, for monitoring of  
 CC anti-HML.24 antibody immunotherapy, and for assay of the antibody or  
 CC serum, urine, milk, synovial fluid or microorganism culture media. The  
 CC method is sensitive down to 500 pg/ml antibody.  
 CC  
 SQ Sequence 143 AA:  
 30  
 Query Match 73.1%; Score 649.5; DB 20; Length 143;  
 Best Local Similarity 95.1%; Pred. No. 6e-55;  
 Matches 135; Conservative 1; Mismatches 3; Indels 3; Gaps 1;  
 QY 39 VPLIIFTIRANSEACRDGLRAVMECHRNTHLLQOELTEAOKGFQDVEAQAATCNHTVMA 98  
 DB 5 vplyagt---nseacdglravmecnrvthlllqelteaokgfgdveaqaatcnhtvma 61  
 QY 99 MASLDAERAKOGKVEELEGELTTLNHKLQDASAEVERLRRENOVLSVRIADKKYPPSSQ 158  
 DB 62 masldeakagqkveelegeeltlnhklqdasaeverlrrenovlsvriadkkypssq 121  
 QY 159 DSSSAAAPQLLIVLIGLSAIIQ 180  
 DB 122 dsssaapqllivlignlsaiiq 143

Search completed: July 18, 2002, 18:43:35  
 Job time: 213 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 18:42:17 ; Search time 27.05 Seconds  
(without alignments)  
1151.168 Million cell updates/sec

Title: US-09-828-217-1  
Perfect score: 889  
Sequence: 1 MASTSYDYGCRPMEDGDKRC.....SSAAPQLLIVLLGLSALLQ 180

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: SP archaea:\*
  - 2: SP bacteria:\*
  - 3: SP fungi:\*
  - 4: SP human:\*
  - 5: SP invertebrate:\*
  - 6: SP mammal:\*
  - 7: SP mhc:\*
  - 8: SP organelle:\*
  - 9: SP phage:\*
  - 10: SP plant:\*
  - 11: SP rodent:\*
  - 12: SP virus:\*
  - 13: SP vertebrate:\*
  - 14: SP unclassified:\*
  - 15: SP virus:\*
  - 16: SP bacteriap:\*
  - 17: SP archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	111	12.5	778	12 O9YMP5	O9YMP5 lymantria d
2	107.5	12.1	461	16 O9K6X4	O9K6X4 bacillus ha
3	107	12.0	143	16 O9KA43	O9KA43 bacillus ha
4	104.5	11.8	756	4 O9Y6W2	O9Y6W2 homo sapien
5	104.5	11.8	756	4 O9NRK8	O9NRK8 homo sapien
6	104.5	11.8	756	4 O9Y6W1	O9Y6W1 homo sapien
7	103.5	11.6	782	4 O9NKK3	O9NKK3 homo sapien
8	103.5	11.6	782	4 O9NXJ4	O9NXJ4 homo sapien
9	102.5	11.5	1534	6 O28298	O28298 canis famli
10	102.5	11.5	1627	5 O96200	O96200 gliardia lam
11	101.5	11.4	304	11 O91YT3	O91YT3 mus musculu
12	101.5	11.4	826	13 O9YHD5	O9YHD5 rana catesb
13	100.5	11.3	506	4 O9NWT9	O9NWT9 homo sapien
14	100	11.2	865	13 O98SN6	O98SN6 gallus gall
15	99.5	11.2	270	6 P79391	P79391 bos taurus
16	99.5	11.2	708	13 O9YHD7	O9YHD7 rana catesb

17	99.5	11.2	1932	13 O98TQ4	O98TQ4 notothenia
18	99	11.1	415	2 O54859	O54859 streptococc
19	99	11.1	800	3 O96X03	O96X03 emeticocella
20	98.5	11.1	712	13 O98SN5	O98SN5 gallus gall
21	98.5	11.1	975	13 O98TQ5	O98TQ5 notothenia
22	98.5	11.1	1931	13 O42352	O42352 cyprinus ca
23	98.5	11.1	1936	13 O90YF6	O90YF6 paracitirhit
24	98.5	11.1	1963	5 O02244	O02244 caenorhabdi
25	97.5	11.0	1938	6 O9GYP9	O9GYP9 oryctolagus
26	97	10.9	425	2 P95808	P95808 streptococc
27	97	10.9	1937	6 O9TV62	O9TV62 sus scrofa
28	97	10.9	1941	5 O26079	O26079 piacopecten
29	97	10.9	1950	5 O26080	O26080 piacopecten
30	97	10.9	2138	5 O9XZ83	O9XZ83 amoeba prot
31	96.5	10.9	1933	13 O90337	O90337 cyprinus ca
32	96	10.8	761	13 P79793	P79793 gallus gall
33	95.5	10.7	173	16 O55887	O55887 synechocyst
34	95.5	10.7	360	2 P70906	P70906 borrelia he
35	95.5	10.7	671	13 O9YHD4	O9YHD4 rana catesb
36	95.5	10.7	977	4 O75300	O75300 homo sapien
37	95.5	10.7	977	4 O9H476	O9H476 homo sapien
38	95.5	10.7	1341	12 O88304	O88304 sandfly fev
39	95.5	10.7	1407	4 O96S82	O96S82 homo sapien
40	95.5	10.7	1586	4 O9P2E9	O9P2E9 homo sapien
41	94.5	10.6	209	13 O9PYE1	O9PYE1 brachydanio
42	94.5	10.6	682	10 O94GH0	O94GH0 oryza sativ
43	94.5	10.6	870	11 O9CS20	O9CS20 mus musculu
44	94.5	10.6	1092	13 O90338	O90338 cyprinus ca
45	94.5	10.6	1388	11 P70336	P70336 mus musculu

ALIGNMENTS

RESULT 1

O9YMP5 PRELIMINARY; PRT; 778 AA.

ID O9YMP5

AC O9YMP5; 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE IDORF-82 PEPTIDE.

OC Lymantria dispar multicaud nuclear polyhedrosis virus (LdMNPV).

OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;

CC Nucleopolyhedrovirus.

OX NCBI\_TaxID=10449;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99124785; PubMed=9887315;

RA Kuzio J., Pearson M.N., Hatwood S.H., Funk C.J., Evans J.T., Slavicek J.M., Rohmann G.F.;

RT "Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria dispar.";

RL Virology 253:17-34(1999).

RN [2]

RP SEQUENCE FROM N.A.

RA Kuzio J.;

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF081810; AAC70268.1; -

DR InterPro; IPR000221; Protamine.P1.

DR Pfam; PF00260; Protamine.P1; 1.

SO SEQUENCE 778 AA; 87841 MW; 811B8BD172CE9E43 CRC64;

Query Match 12.5%; Score 111; DB 12; Length 778;  
Best Local Similarity 32.5%; Pred. No. 0.4;  
Matches 38; Conservative 22; Mismatches 47; Indels 10; Gaps 4;

OY 46 IKANSEACRDGIRAYMECR-NWTHILLOEITFAQKGFQDVEAOATCNITVMAIASIDA 104  
DB 520 IKAQSELNMD-LQAKAEAOADANARLQAEISLKR---AESDADLRRNRVAQLAEAES 574  
OY 105 EKAQCGKKEVELEGEITTLNHLQDASAEVERLRENGVLSVRIDKRYYPSSQSS 161

DB 575 GAADLQNRIRARLEAESGILTRLOESAAEVLALRROKEDLERRTAAS----AAQDVS 627

RESULT 2

ID 09K6X4 PRELIMINARY; PRT; 461 AA.

AC 09K6X4;

DT 01-OCT-2000 (TREMblrel. 15, Created)

DT 01-OCT-2000 (TREMblrel. 15, last sequence update)

DT 01-DEC-2001 (TREMblrel. 19, last annotation update)

DE CELL WALL-BINDING PROTEIN.

GN BH3600.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI\_TaxID=86665;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=C-125 / JCM 9153;

RA MEDLINE=20512582; PubMed=11058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.;

RT "Complete genome sequence of the alkaliphilic bacterium Bacillus

RT halodurans and genomic sequence comparison with Bacillus subtilis."

RL Nucleic Acids Res. 28:4317-4331(2000).

DR EMBL: AP001519; BAB07319.1; -

DR MEROPS: M37.UPW; -

DR InterPro: IPR002886; Peptidase\_M37.

DR Pfam: PF01551; Peptidase\_M37; 1.

KW Complete proteome.

SQ SEQUENCE 461 AA; 50372 MW; 2918480CD67AF3F CRC64;

Query Match 12.1%; Score 107.5; DB 16; Length 461;

Best Local Similarity 26.1%; Pred. No. 0.44;

Matches 31; Conservative 28; Mismatches 55; Indels 5; Gaps 2;

QY 33 IIVILGVLIIITIKANSEACRDGLRAVMECRVTHLLQOELTEAQKGFQDVEQAATCN 92

DB 5 ISLVAAAGLTFILFSQSSIEDA-KANSSLOQISDVQKREKQEKTELEL----- 59

QY 93 HTYMALMASIDAKAQQKQKVEELGEGITTLNHLQDASAEVERLRRENOVLSVRIADK 151

DB 60 KEVEKEIGDITAEIERIDKEVETSGKIOEKREIEVEVQAEIIELEKQELIERIAER 118

SQ

RESULT 3

ID 09KA43 PRELIMINARY; PRT; 143 AA.

AC 09KA43;

DT 01-OCT-2000 (TREMblrel. 15, Created)

DT 01-OCT-2000 (TREMblrel. 15, last sequence update)

DT 01-JUN-2001 (TREMblrel. 17, last annotation update)

DE FLAGELLAR PROTEIN REQUIRED FOR FLAGELLAR FORMATION.

GN FLIL OR BH2447.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI\_TaxID=86665;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=C-125 / JCM 9153;

RA MEDLINE=20512582; PubMed=11058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.;

RT "Complete genome sequence of the alkaliphilic bacterium Bacillus

RT halodurans and genomic sequence comparison with Bacillus subtilis."

RL Nucleic Acids Res. 28:4317-4331(2000).

DR EMBL: AP001515; BAB06166.1; -

KW Flagella; Complete proteome.

SQ SEQUENCE 143 AA; 16386 MW; 613446064B1C5402 CRC64;

Query Match 12.0%; Score 107; DB 16; Length 143;

Best Local Similarity 22.1%; Pred. No. 0.13;

Matches 31; Conservative 34; Mismatches 49; Indels 26; Gaps 4;

QY 23 LIGIGIIVLLIIVILGVLIIITIKANSEACRDGLRAVMEC-----RNVTHLLQOEL 74

DB 6 LVNIMLIIIVLIVLGVAVLIFVNYFNNEDEQDREPTIDEIIAQSYTEIEITNLLSNDP 65

QY 75 TEA-----QKGFQDVEQAATCNHTVMAIASIDAQAQKQKVEELGEGITTLNHLK 126

DB 66 VRAEFLIHVDNRNALDEVQRDFQVNNIIIRSLAGDASQLSGADGTEKLEAO----- 118

QY 127 LODASAEVERLRRENOVLSV 146

DB 119 LQD---DINLMOEGSVYKI 135

RESULT 4

ID 09Y6W2 PRELIMINARY; PRT; 756 AA.

AC 09Y6W2;

DT 01-NOV-1999 (TREMblrel. 12, Created)

DT 01-NOV-1999 (TREMblrel. 12, last sequence update)

DT 01-DEC-2001 (TREMblrel. 19, last annotation update)

DE HCR PROTEIN.

GN HCR.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=20014706; PubMed=10545595;

RA Oka A., Tamiya G., Tomizawa M., Ota M., Katsuyama Y., Makino S.,

RA Shima T., Yoshitome M., Lituka M., Sasaki Y., Iwashita K.,

RA Kawakubo Y., Sugai J., Ozawa A., Ohkido M., Kimura M., Bahram S.,

RA Inoko H.;

RT "Association analysis using refined microsatellite markers localizes a

RT susceptibility locus for psoriasis vulgaris within a 11kb segment

RL Hum. Mol. Genet. 8:2165-2170(1999).

DR EMBL: AB029331; BAA81890.1; -

SQ SEQUENCE 756 AA; 86118 MW; 88FD5F85BEF07601 CRC64;

Query Match 11.8%; Score 104.5; DB 4; Length 756;

Best Local Similarity 24.7%; Pred. No. 1.4;

Matches 39; Conservative 34; Mismatches 48; Indels 37; Gaps 7;

QY 54 RDLGLRAVME-----CRNVTHLL---QOELTEAQKGFQDVEQAQ-AATC-----NHTVMA 97

DB 258 RDSLHMTAEILLQYRVOSLTHLLQOELTRKVPQSDSEPEPTRKQCSILNWRREKVF 317

QY 98 IMAISLDAKAQKQKVEELGEGITTLNHLK-----LODASAEVERLRRENOV 143

DB 318 LMTQAKAQELHSDSVKQLKGVAISLOEKVTSQSQEQALIQRSLOKRAAEVERBGAAGK 377

QY 144 LSVRIA-----DKKTYFPSSQDSSAAPQILLIYLGLISA 177

DB 378 LOELSRAGARRW---QOOTASAEQILRVNAVVS 412

RESULT 5

ID 09NRK8 PRELIMINARY; PRT; 756 AA.

AC 09NRK8;

DT 01-OCT-2000 (TREMblrel. 15, Created)

DT 01-OCT-2000 (TREMblrel. 15, last sequence update)

DT 01-DEC-2001 (TREMblrel. 19, last annotation update)

DE A-HELICAL PROTEIN.

GN HCR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20347693; PubMed=10888604;  
RA Asunuma K., Iitonen T., Iitonen V., Lokki M.-L.,  
RA Suomalainen S., Snellman E., Saarialho-Kere U., Kere J.,  
RT "A candidate gene for psoriasis near HLA-C, HCR (Pg8), is highly  
RT polymorphic with a disease-associated susceptibility allele."  
RL Hum. Mol. Genet. 9:1533-1542(2000).  
DR EMBL: AF216493; AAF74221.1;  
SQ SEQUENCE 756 AA; 86058 MW; BE4D03358B62DEB4 CRC64;

Query Match 11.8%; Score 104.5; DB 4; Length 756;  
Best Local Similarity 24.7%; Pred. No. 1.4;  
Matches 39; Conservative 34; Mismatches 48; Indels 37; Gaps 7;

54 RDGLRAVME-----CRNVTLL-----QDELTEAKGFQDVEAO-AATC-----NHTYMA 97  
Db 258 RDSLHATFAELLQVRVQSLTHIALQEEELTRKVPSPDLEPFRKCCSLNRMREKVFA 317  
OY 98 LMSLDIAKAGOKKVELEGEITTLNKK-----LQDASAEVERLRENOV 143  
Db 318 LMSLDIAKAGOKKVELEGEITTLNKK-----LQDASAEVERLRENOV 143  
OY 144 LSVRIA-----DKKYPSSODSSSAAPOLLIVLLGSA 177  
Db 378 LQELSLRAQAEARRRW-----QQQTASAEQLRLVYNAVSS 412

RESULT 6  
OY6W1 PRELIMINARY; PRT; 756 AA.  
AC OY6W1;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HCR PROTEIN (TRICOMYALIN HOMOLOGUE).  
GN HCR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RT TISSUE=BLOOD;  
RA Oka A., Tamaya G., Makino S., Tomizawa M., Yamagata T., Shilina T.,  
RA Watanabe K., Yamazaki M., Tashiro H., Kimura M., Inoko H.,  
RT "HCR-a-helical coiled-coil rod homologue."  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Hirakawa M., Yamaguchi H., Imai K., Shimada J.,  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Shilina S., Tamaya G., Oka A., Inoko H.,  
RT "Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region."  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB029343; BAA82158.1;  
DR EMBL: AP000509; BAB63313.1;  
SQ SEQUENCE 756 AA; 85951 MW; 3D109AFOEEFAB9F5 CRC64;

Query Match 11.8%; Score 104.5; DB 4; Length 756;  
Best Local Similarity 25.2%; Pred. No. 1.4;  
Matches 40; Conservative 32; Mismatches 48; Indels 39; Gaps 7;  
OY 54 RDGLRAVME-----CRNVTLL-----QDELTEAKGFQDVEAO-AATC-----NHTYMA 97

Db 258 RDSLHATFAELLQVRVQSLTHIALQEEELTRKVPSPDLEPFRKCCSLNRMREKVFA 317  
OY 98 LMSLDIAKAGOKKVELEGEITTLNKK-----LQDASAEVERLRENOV 143  
Db 318 LMSLDIAKAGOKKVELEGEITTLNKK-----LQDASAEVERLRENOV 143  
OY 144 LSVRIA-----DKKYPSSODSSSAAPOLLIVLLGSA 177  
Db 378 LQELSLRAQAEARRRW-----QQQTASAEQLRLVYNAVSS 412

RESULT 7  
OY6W1 PRELIMINARY; PRT; 782 AA.  
AC OY6W1;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE CDNA FLJ20197 FIS, CLONE COLF0996.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RT TISSUE=COLON MUCOSA;  
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,  
RA Nakamura Y., Isono S., Sugano S.,  
RT "NEDD human CDNA sequencing project."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AK000204; BAA91007.1;  
SQ SEQUENCE 782 AA; 86616 MW; F472FE544F627CE8 CRC64;

Query Match 11.6%; Score 103.5; DB 4; Length 782;  
Best Local Similarity 25.2%; Pred. No. 1.7;  
Matches 40; Conservative 32; Mismatches 48; Indels 39; Gaps 7;  
OY 54 RDGLRAVME-----CRNVTLL-----QDELTEAKGFQDVEAO-AATC-----NHTYMA 97  
Db 284 RDSLHATFAELLQVRVQSLTHIALQEEELTRKVPSPDLEPFRKCCSLNRMREKVFA 343  
OY 98 LMSLDIAKAGOKKVELEGEITTLNKK-----LQDASAEVERLRENOV 143  
Db 344 LMSLDIAKAGOKKVELEGEITTLNKK-----LQDASAEVERLRENOV 143  
OY 144 LSVRIA-----DKKYPSSODSSSAAPOLLIVLLGSA 177  
Db 404 LQELSLRAQAEARRRW-----QQQTASAEQLRLVYNAVSS 438

RESULT 8  
OY6W1 PRELIMINARY; PRT; 782 AA.  
AC OY6W1;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE CDNA FLJ20210 FIS, CLONE COLF1787.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RT TISSUE=COLON MUCOSA;  
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,  
RA Nakamura Y., Isono S., Sugano S.,  
RT "NEDD human CDNA sequencing project."  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AK000217; BAA91016.1; -  
SQ SEQUENCE 782 AA; 88643 MW; A510C8BCFA8247B CRC64;

Query Match  
Best Local Similarity 25.2%; Pred. No. 1.7;  
Matches 40; Conservative 32; Mismatches 48; Indels 39; Gaps 7;

QY 54 RDGLRAVME-----CRNVTHL---OOELTEAOKGFQDVEAQAATCNHTVMALMSL-DA 104  
DB 284 RDSLHATTAELLQVRVOSLTHIALOEELTRKYVQPSDSEPEPTRKQCSLNNRREKVFA 343  
QY 98 LMSLDAEKAQOKKVEELEGETTINHK-----LQDASAEVELRRENOY 143  
DB 344 LMVQLAKAELEHSDVKQKLGQVASLOEKVTSQSOQAILQRSLODKAAEVEVERGAKG 403  
QY 144 LSVRI-----ADKKYPPSSODSSSAAPOLLIVLGLISA 177  
DB 404 LQELSLRAQEARROW-----QQQTASAEQLRLVYNAVSS 438

RESULT 9

ID 028298 PRELIMINARY; PRT; 1534 AA.  
AC 028298;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE RIBOSOME RECEPTOR.  
GN P180.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95310363; PubMed=7790375;  
RA Wanner E.E., Sun Y., Savitz A.J., Meyer D.I.;  
RT "Functional characterization of the 180 kDa ribosome receptor in vivo."  
RL J. Cell Biol. 130:29-39(1995).  
DR EMBL: X87224; CAA60676.1; -  
KW Receptor.  
SQ SEQUENCE 1534 AA; 164586 MW; B343BCF12656F3C5 CRC64;

Query Match  
Best Local Similarity 11.5%; Score 102.5; DB 6; Length 1534;

Matches 26; Conservative 22; Mismatches 42; Indels 1; Gaps 1;

QY 55 DGLRAVMECRNVTHLLOELTEAOKGFQDVEAQAATCNHTVMALMSLDAEKAQOKKVE 114  
DB 847 DAAVAASKIRLEVKKELAAEKAAGAKYKQLVAREGETFVQARIETASYREHYKEVQ 906  
QY 115 ELEGETTINHKIQDA-SAEVELRRENOY 144  
DB 907 QLOGKIRTLQEOLENGPNTOLARLQOENSIL 937

RESULT 10

ID 096200 PRELIMINARY; PRT; 1627 AA.  
AC 096200;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE AXONEME-ASSOCIATED PROTEIN GASP-180.  
OS Giardia lamblia (Giardia intestinalis).  
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinales; Giardia.  
OX NCBI\_TaxID=5741;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Elmendorf H.G., Rohrer S.C., La Vigne E.A., Nash T.E.;

RT "Novel Axoneme-Associated Proteins in Giardia lamblia."  
RL Submitted (Jul-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF400249; AAK91740.1; -  
SQ SEQUENCE 1627 AA; 179270 MW; 021EED9763907DCC CRC64;

Query Match  
Best Local Similarity 11.5%; Score 102.5; DB 5; Length 1627;  
Matches 34; Conservative 23; Mismatches 60; Indels 5; Gaps 3;

QY 46 IRANSACRDGLRAVMECRNVTHLLOELTEAOKGFQDVEAQAATCNHTVMALMSL-DA 104  
DB 1284 LRESAPALQDKLHAUSDSDRADBDLOKLYEQLKDGAKELVAERDATTDELKQRLRPT 1343  
QY 105 EKAQOKK-RVEELEGETTINHKIQDASAEVELRRENOYLSRIADKKYPPSSODSSSA 163  
DB 1344 EEVDLKEKIELDELDEIVNLNGLKKDKDELIELRLRQ---LEAQPTATVYVESGSEVGD 1400  
QY 164 AA 165  
DB 1401 AA 1402

RESULT 11

ID 091Y73 PRELIMINARY; PRT; 304 AA.  
AC 091Y73;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE MACROPHAGE GALACTOSE N-ACETYL-GALACTOSAMINE SPECIFIC LECTIN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC014811; AAH14811.1; -  
KW Lectin.  
SQ SEQUENCE 304 AA; 34622 MW; 2272E1ADA2C0262A CRC64;

Query Match  
Best Local Similarity 11.4%; Score 101.5; DB 11; Length 304;  
Matches 36; Conservative 38; Mismatches 48; Indels 33; Gaps 7;

QY 24 LGIGILVLIIVILGVPLIITIKANSEACRD-GIRAVMECRNVTHLLOELTEAOKGF 81  
DB 41 LGISLILVVSIVIG-----SQNSQLRBDLGTIRALID--NTTSKIKAE-----F 83  
QY 82 GQVEAQAATCNHTVMALMSLDAEKAQO-----KQVELEGET---TTINHKIQDAS 131  
DB 84 QSLDSRAUDSEFGISLKYDVEDHROELOAGRDLSOKYTSLESTLEKRPQALKTLDLSTL 143  
QY 132 AEVELRRENOYLSRIADKKYPPSSODSSSAAP 166  
DB 144 DHVQQLRKDLKALTOQLANLK-----NNGSEVACCP 174

RESULT 12

ID 09YHD5 PRELIMINARY; PRT; 826 AA.  
AC 09YHD5;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE MYOSIN HEAVY CHAIN (FRAGMENT).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranidae; Rana.

[illegible]

ID		PRT:	865 AA.
AC	098SN6;		
DT	01-JUN-2001 (TREMBlrel. 17, Created)		
DR	01-JUN-2001 (TREMBlrel. 17, Last sequence update)		
DE	01-DEC-2001 (TREMBlrel. 19, Last annotation update)		
OE	RHO-ASSOCIATED COILED-COIL FORMING KINASE 1 (FRAGMENT).		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Archosauromia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OX	NCHI_TaxID=9031;		
RP	[1]		
RA	SEQUENCE FROM N.A.		
RF	Wei L., Roberts W., Wang L., Yamada M., Zhang S., Zhao Z.,		
RT	Rivkees S.A., Schwartz R.J., Imanaka-Yoshida K.;		
RL	"Rho kinases play an obligatory role in vertebrate embryonic organogenesis.";		
CC	Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases. -I- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. EMBL: AF347075; AAC29627.1; - HSSP: P05132; IAPM. InterPro: IPRO00719; Euk_pkinase. InterPro: IPRO00961; Pkinase.C. InterPro: IPRO00861; REM_repeat. InterPro: IPRO02290; Ser_thr_pkinase. InterPro: IPRO01245; Tyr_pkinase. Pfam: PF02185; HRI; 1. Pfam: PF00069; pkinase; 1. PRINTS: PR00109; TYRKINASE. SMART: SMART0074; HRI; 1. SMART: SMART0220; S_TRK; 1. SMART: SMART0133; S_TRK_X; 1. SMART: SMART0219; TyrcK; 1. PROSITE: PS00107; PROTEIN_KINASE_ATP; 1. PROSITE: PS0011; PROTEIN_KINASE_DOM; 1. PROSITE: PS00108; PROTEIN_KINASE_ST; 1. ATP-binding; kinase; Serine/threonine-protein kinase; Transferase. FT NON_TER FT NO_TER SQ SEQUENCE 865 AA; 101138 MW; 2ADIA3AD7CC8352A CRC64;		
Query Match	11.2%; Score 100; DB 13; Length 865;		
Best Local Similarity	24.3%; Pred. No. 3.8;		
Matches	33; Conservative 30; Mismatches 47; Indels 26; Gaps 5;		
OY	46 IKANSEACRDGIGRAVMECRNVTYLHQEETFEOKGFQOVEAA-----ATCN 92		
DB	732 LKGSERKOLKEINTLLLEAK---RLLEVLAQLAKYRGNEGOMRLQDLAEQYFTSLY 788		
OY	93 HT-VAMLASDAEKAGQOKVEELGETTLTNLRKIQ--DASAEVLRLRENVLSRIA 149		
DB	789 KTQVVELKEEDIEKKNKETORKMOELONEKETITLTOLDIAETKAESSEQAL-----ALL 841		
OY	150 DKKYPPSSODSSSAQA 165		
DB	842 EEQYFELSSESRRKAS 857		
RESULT	15		
ID	P79391		
AC	P79391;		
DT	01-MAY-1997 (TREMBLrel. 03, Created)		
DR	01-MAY-1997 (TREMBLrel. 03, Last sequence update)		
DE	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
OE	LECTIN-LIKE OXIDIZED LDL RECEPTOR.		
OS	Bos taurus (bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;		



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2002, 18:42:42 ; Search time 11.94 Seconds  
(without alignments)  
583.712 Million cell updates/sec

Title: US-09-828-217-1

Perfect score: 889  
Sequence: 1 MASTSYDYCRVPMEDGDKRC.....SSAAPQLLVLIGLSALLQ 180

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match length	ID	Description
1	889	100.0	180	BST2_HUMAN	Q10589 homo sapien
2	102	11.5	550	KUCR_RAT	P10716 rattus norv
3	101	11.4	422	KICR_MOUSE	P49384 mus musculu
4	99.5	11.2	304	MUGL_MOUSE	O60763 homo sapien
5	99	11.1	962	VDP_HUMAN	P02566 caenorhabdi
6	98.5	11.1	1966	MYB_CAEEL	Q28641 oryctolagus
7	98	11.0	1938	MYH4_RABIT	P02567 caenorhabdi
8	98	11.0	1938	MYSD_CAEEL	P22312 sclara copr
9	97	10.9	286	PUR2_SCICO	P41542 rattus norv
10	97	10.9	959	VDP_RAT	P34531 caenorhabdi
11	96.5	10.9	893	VM92_CAEEL	P22311 sclara copr
12	96	10.8	286	PUR1_SCICO	P13535 homo sapien
13	95	10.7	1937	MYH8_HUMAN	P35748 oryctolagus
14	95	10.7	1972	MYHB_RABIT	P35748 oryctolagus
15	95	10.7	1972	MYHB_HUMAN	P35748 oryctolagus
16	94.5	10.6	1935	MYSS_CYPCA	O90339 cyprinus ca
17	94.5	10.6	4473	PLE1_CRIGR	O91155 citreulius
18	94	10.6	1084	MYSS_RABIT	P02563 oryctolagus
19	93.5	10.5	1938	MYH6_RAT	P02563 rattus norv
20	93.5	10.5	1939	MYH6_MESAU	P13539 mesocricetu
21	93	10.5	848	MYSP_DIRIM	P13592 difrofilaria
22	93	10.5	879	MYSP_ONCVO	O02171 onchocerca
23	93	10.5	880	MYSP_BRUMA	O01202 brugia mala
24	93	10.5	1972	MYHB_MOUSE	O08638 mus musculu
25	92.5	10.4	1509	MYSN_ACACA	P02566 mus musculu
26	92.5	10.4	1938	MYH6_MOUSE	O02566 mus musculu
27	92	10.3	882	MYSP_CAEEL	P10567 caenorhabdi
28	91.5	10.3	244	MYH7_PAPHA	P11778 papio hamad
29	91.5	10.3	428	PLT2_CARAU	O42305 carassius a
30	91.5	10.3	1102	MYSC_CHICK	P26616 gallus galli
31	91.5	10.3	1934	MYH7_MESAU	P13540 mesocricetu
32	91.5	10.3	1935	MYH7_HUMAN	P12883 homo sapien
33	91.5	10.3	1935	MYH7_RAT	P02564 rattus norv

34	91	10.2	407	1	M21_STRPY	P50468 streptococc
35	91	10.2	516	1	P54_ENTEC	P13692 enterococcu
36	90.5	10.2	692	1	MYS_PODCA	O05000 podocoryne
37	90.5	10.2	1846	1	MYB_RAT	P70563 rattus norv
38	90.5	10.2	1938	1	MYSS_CHICK	P13538 gallus galli
39	90.5	10.2	1939	1	MYH1_HUMAN	P12882 homo sapien
40	90.5	10.2	4687	1	MYB_MOUSE	P30427 rattus norv
41	90	10.1	941	1	VDP_MOUSE	O92120 mus musculu
42	90	10.1	2116	1	MYSD_DICDI	P08799 dictyosteli
43	89.5	10.1	1938	1	MYHD_HUMAN	Q9ukx3 homo sapien
44	89.5	10.1	1939	1	MYHB_HUMAN	P13533 homo sapien
45	89	10.0	1978	1	MYHB_CHICK	P10587 gallus galli

## ALIGNMENTS

RESULT	ID	STANDARD	PRT	180 AA.
BST2_HUMAN	Q10589			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Bone marrow stromal antigen 2 (BST-2).			
GN	BST2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=95331788; PubMed=7607676;			
RA	Ishtikawa J., Kaisho T., Tomizawa H., Lee B.O., Kobune Y.,			
RA	Inazawa J., Oritani K., Itoh M., Ochi T., Ishihara K., Hirano T.;			
RT	"Molecular cloning and chromosomal mapping of a bone marrow stromal			
RT	cell surface gene, BST2, that may be involved in pre-B-cell growth.";			
RL	Genomics 26:527-534(1995).			
CC	- FUNCTION: MAY BE INVOLVED IN PRE-B-CELL GROWTH.			
CC	- SUBCELLULAR LOCATION: Type II membrane protein.			
CC	- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LIVER, HEART			
CC	AND PLACENTA. LOWER LEVELS IN PANCREAS, KIDNEY, SKELETAL MUSCLE			
CC	AND BRAIN.			
CC	- DISEASE: MAY PLAY A ROLE IN B-CELL ACTIVATION IN RHEUMATOID			
CC	ARTHRITIS (RA).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL: D28137; BAA05679.1; -			
DR	MIM: 600534; -			
KM	Transmembrane: Glycoprotein; Signal-anchor; Polymorphism.			
FT	DOMAIN 1 20			
FT	TRANSMEM			
FT	DOMAIN 49 180			
FT	CARBOHYD			
FT	CARBOHYD 92 92			
FT	VARIANT 143 143			
FT	SEQUENCE 180 AA; 19769 MW; CAF52340D69061EE CRC64;			
QY	1 MASTSYDYCRVPMEDGDKRCLLIGILVLLIYILVPLITFTIKNSACRGRLRAY 60			
QY				

DB 1 MASTSDYCVPMEDGDKRCKLLGIGLIVLIVLIGVPLIIFTIKANSEACRDLIRAV 60  
 QY 61 MECRNVTHLLQOELTEAOKGFQDVEAQAATCNHTVMAIAMSILDAEKAQOGKVEELEGEL 120  
 DB 61 MECRNVTHLLQOELTEAOKGFQDVEAQAATCNHTVMAIAMSILDAEKAQOGKVEELEGEL 120  
 QY 121 TTNHKLQDASAVEERLRRENOVLSVRIDAKKRYPPSSQDSSAAAPQLILVILGLSLAQ 180  
 DB 121 TTNHKLQDASAVEERLRRENOVLSVRIDAKKRYPPSSQDSSAAAPQLILVILGLSLAQ 180  
 RESULT 2  
 KICR\_RAT STANDARD: PRT: 550 AA.  
 AC P10716;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 RT Kupffer cell receptor.  
 KCLR.  
 US Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 83-104.  
 RX MEDLINE=88227939; PubMed=2836387;  
 RA Hoyle G.W., Hill R.L.;  
 RT "Molecular cloning and sequencing of a cDNA for a carbohydrate  
 binding receptor unique to rat Kupffer cells.";  
 RT J. Biol. Chem. 263:7487-7492(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91107689; PubMed=1846367;  
 RA Hoyle G.W., Hill R.L.;  
 RT "Structure of the gene for a carbohydrate-binding receptor unique to  
 rat Kupffer cells.";  
 RT J. Biol. Chem. 266:1850-1857(1991).  
 CC -1- FUNCTION: RECEPTOR WITH AN AFFINITY FOR GALACTOSE AND FUCOSE.  
 CC COULD BE INVOLVED IN ENDOCYTOSIS.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein.  
 CC -1- TISSUE SPECIFICITY: KUPFER CELLS.  
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL J03734; AAA41472.1; -  
 DR EMBL M55532; AAA40892.1; -  
 DR PIR A28166; A28166.  
 DR PIR A38674; A38674.  
 DR HSP; P20693; 1HLJ.  
 DR InterPro: IPR000017; Syntaxin.  
 DR InterPro: IPR001304; lectin\_c.  
 DR Pfam: PF00059; lectin\_c; 1.  
 DR SMART; SM00034; CLECT; 1.  
 DR SMART; SM00503; SYN; 1.  
 DR PROSITE; PS00615; C-TYPE\_LECTIN\_1; 1.  
 DR PROSITE; PS50041; C-TYPE\_LECTIN\_2; 1.  
 KM Receptor; Transmembrane; Glycoprotein; Lectin; Signal-anchor;  
 KM Endocytosis.  
 FT DOMAIN 1 42  
 FT TRANSMEM 43 69  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 70 550  
 FT DOMAIN 438 538  
 FT DISULFID 440 536  
 FT DISULFID 516 528  
 FT BY SIMILARITY.  
 FT BY SIMILARITY.

FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 550 AA; 61104 MW; 9358A6CF4C306270 CRC64;  
 Query Match 11.5%; Score 102; DB 1; Length 550;  
 Best local Similarity 28.0%; Pred. No. 0.57;  
 Matches 37; Conservative 20; Mismatches 61; Indels 14; Gaps 3;  
 QY 46 IKANSEACRDLGRAVMECRNVTHLLQOELTEAOKGFQDVEAQAATCNHTVMAIAMSILDAE 105  
 DB 256 ISAEIOAMDQMDQORAGE-----EWTSLKDELTTLTAQIONANGHLEQDTQIQGL 305  
 QY 106 KAQCKQKVEELEGELTTLNHKLQDASAVEERLRRENOVLSVRIDAKKRYPPSSQDSSAA 165  
 DB 306 KAO-LKSTSLNSQILEVNGKLDSSRELQTLRD--LSDVSAKSNVQMLQSNLQAK 361  
 QY 166 POLIIVILGLSA 177  
 DB 362 AEVQSLKTGLEA 373  
 RESULT 3  
 KICR\_MOUSE STANDARD: PRT: 422 AA.  
 AC P05784; O61766;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Keratin, type I cytoskeletal 18 (Cytokeratin 18) (Cytokeratin endo B)  
 DE (Keratin D).  
 GN KRT18 OR KRT1-18 OR KERD.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89196920; PubMed=2467843;  
 RA Ichinose Y., Morita T., Zhang F., Srinivasanongram S., Tondella M.L.C.,  
 RA Matsumoto M., Nozaki M., Matsushiro A.;  
 RT "Nucleotide sequence and structure of the mouse cytokeatin endo B  
 gene.";  
 RT Gene 70:85-95(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Teratocarcinoma;  
 RA Alonso A., Weber T., Jorcano J.L.;  
 RT "Cloning and characterization of keratin D, a murine endodermal  
 cytoskeletal protein induced during in vitro differentiation of F9  
 teratocarcinoma cells.";  
 RT Roux's Arch. Dev. Biol. 196:16-21(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86085876; PubMed=2416755;  
 RA Singer P.A., Trevor K., Oshima R.G.;  
 RT "Molecular cloning and characterization of the Endo B cytokeatin  
 expressed in preimplantation mouse embryos.";  
 RT J. Biol. Chem. 261:538-547(1986).  
 RN [4]  
 RP SEQUENCE OF 1-131 FROM N.A.  
 RX MEDLINE=8825838; PubMed=2454868;  
 RA Oshima R.G., Trevor K., Shevinsky L.H., Ryder O.A., Cecena G.;  
 RT "Identification of the gene coding for the Endo B murine cytokeatin  
 and its methylated, stable inactive state in mouse nonepithelial  
 cells.";  
 RT Genes Dev. 2:505-516(1988).  
 CC -1- SUBUNIT: HETEROTRIMER OF TWO TYPE I AND TWO TYPE II KERATINS.  
 CC KERATIN 18 ASSOCIATES WITH KERATIN 8.



CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND  
CC MICROFILAMENT KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC)  
CC (40-55 AND 56-70 KILODALTONS, RESPECTIVELY).  
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: M22832; AAA37552.1; -  
DR EMBL: M36376; AAA39373.1; -  
DR EMBL: M11686; AAA39390.1; -  
DR EMBL: Y00217; CAA68365.1; -  
DR PIR: A25621; A25621.  
DR PIR: A28428; A28428.  
DR PIR: J70406; J70406.  
DR SWISS-2DPAGE: P05784; MOUSE.  
DR MGI: 96692; Krt1-18.  
DR InterPro: IPR001664; IF.  
DR InterPro: IPR002957; Keratin\_I.  
DR Pfam: PF00038; filament; 1.  
DR PRINTS: PR01248; TYPE1KERATIN.  
DR PROSITE: PS00226; IF, 1.  
KW Intermediate filament; Coiled coil; Keratin; Glycoprotein;  
KW Acetylation.  
FT INT\_MET 0 0 BY SIMILARITY.  
FT MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
FT DOMAIN 1 70 HEAD.  
FT DOMAIN 1 71 ROD.  
FT DOMAIN 380 422 TAIL.  
FT DOMAIN 71 106 COIL 1A.  
FT DOMAIN 107 124 LINKER 1.  
FT DOMAIN 125 216 COIL 1B.  
FT DOMAIN 217 240 LINKER 12.  
FT DOMAIN 241 379 COIL 2.  
FT SITE 263 263 STUTTER.  
FT SITE 323 323 STUTTER.  
FT CARBOHYD 30 30 O-LINKED (GLCNAC) (BY SIMILARITY).  
FT CARBOHYD 31 31 O-LINKED (GLCNAC) (BY SIMILARITY).  
FT CARBOHYD 49 49 O-LINKED (GLCNAC) (BY SIMILARITY).  
FT CONFLICT 133 133 L -> F (IN REF. 1).  
FT CONFLICT 243 243 D -> N (IN REF. 2).  
FT CONFLICT 252 252 A -> G (IN REF. 2).  
SQ SEQUENCE 422 AA; 47373 MW; 4D5B0E9C7732FE2F3 CRC64;

Query Match 11.4%; Score 101; DB 1; Length 422;  
Best Local Similarity 27.4%; Pred. No. 0.51;  
Matches 51; Conservative 27; Mismatches 58; Indels 50; Gaps 8;

QY 44 FTIRANSE-ACRD-----GLRAVMECRNVTYHL-----LOQELTEAOKG---POD 83  
Db 156 FRVKTETELARQKQVESDIHGLRKAYVDNTNITRLQLETEIALKEELFMKKNHEEVOG 215  
QY 84 VEAQAATCNHTV-----MALMASLDAE-KAOGOKVVEELE-----GETT 121  
Db 216 LEAQIASSGLIYVEDAPKQSDLSKIMADIRAQVELAKNRELDKYSQSEESTVVT 275  
QY 122 TLNKLQDASAEVERLARENOVLSYRIADKTYPSQSS-----SAAAPQLLIYVL 173  
Db 276 TRKSAIRIAETTLTLELRTIQTLELIDSMKNQINLENISGDVARYKAQMOQLNGVL 335  
QY 174 GLSALL 179  
Db 336 HLESEL 341

RESULT 4  
MGL\_MOUSE

ID MGL\_MOUSE STANDARD; PRT; 304 AA.  
AC P49300;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Macrophage asialoglycoprotein-binding protein (M-ASGP-BP) (Macrophage  
DE galactose/N-acetylgalactosamine-specific lectin) (MGL).  
GN MGL.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C3H/HEN;  
RX MEDLINE=92268032; PubMed=1587794;  
RA Sato M., Kawakami K., Osawa T., Toyoshima S.;  
RT "Molecular cloning and expression of cDNA encoding a galactose/N-  
RT acetylgalactosamine-specific lectin on mouse tumoricidal  
RT macrophages";  
RL J. Biochem. 111:331-336(1992).  
RN [2]  
RP SEQUENCE OF 102-120 AND 137-151.  
RC STRAIN=C3H/HEN;  
RX MEDLINE=89197865; PubMed=3241002;  
RA Oda S., Sato M., Toyoshima S., Osawa T.;  
RT "Purification and characterization of a lectin-like molecule specific  
RT for galactose/N-acetyl-galactosamine from tumoricidal macrophages";  
RL J. Biochem. 104:600-605(1988).  
CC -1- FUNCTION: RECOGNIZES TERMINAL GALACTOSE AND N-ACETYLGALACTOSAMINE  
CC UNITS. MAY PARTICIPATE IN THE INTERACTION BETWEEN TUMORICIDAL  
CC MACROPHAGES AND TUMOR CELLS.  
CC -1- SUBUNIT: HOMO-OLIGOMER.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.  
CC -1- TISSUE SPECIFICITY: IS EXPRESSED ON THE SURFACE OF ACTIVATED  
CC MACROPHAGES.  
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: S36676; AAB22171.1; -  
DR HSSP: P06734; IKJE.  
DR MGI: 96975; MGL.  
DR InterPro: IPR001304; lectin\_c.  
DR Pfam: PF00059; lectin\_c; 1.  
DR SMART: SM00034; CLECT; 1.  
DR PROSITE: PS00615; C-TYPE LECTIN\_1; 1.  
DR PROSITE: PS50041; C-TYPE LECTIN\_2; 1.  
KW Lectin; Glycoprotein; Transmembrane; Calcium; Signal-anchor.  
FT DOMAIN 1 35 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 36 56 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
FT (POTENTIAL).  
FT DOMAIN 57 304 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 172 298 C-TYPE LECTIN (LONG FORM).  
FT DISULFD 173 184 BY SIMILARITY.  
FT DISULFD 201 296 BY SIMILARITY.  
FT DISULFD 274 288 BY SIMILARITY.  
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 304 AA; 34596 MW; 3F79CD12C34F5BCC CRC64;

Query Match 11.2%; Score 99.5; DB 1; Length 304;  
Best Local Similarity 23.2%; Pred. No. 0.47;  
Matches 36; Conservative 37; Mismatches 49; Indels 33; Gaps 7;

QY 24 LGIGLVLLIIVLIGVPLIIFIRKANSEACRD--GLRAVMECRNVTYHLQQLTEAOKGF 81

Db 41 LGSLLLLVVSVIG-----SQNSQLRRDGLTTLATLD--NTTSKIAE-----F 83  
 QY 82 QDVEAQAATCNHTVMALMASLDAEKAAGQ-----KKVELEGEI-----TTLNKKLQDAS 131  
 Db 84 QSDSRADSEKKGSSKAYVEDHROELQAGRDLSOKVTSLESTVEKREQALKTDLSDLT 143  
 QY 132 AEVERLRENOVLVRIADKKYPPSSODSSSAAP 166  
 Db 144 DHVQRLKRDKLALCOLANLK-----NNGSEVACCP 174

## RESULT 5

VDP\_HUMAN STANDARD; PRT; 962 AA.

AC 060763;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE General vesicular transport factor p115 (transcytosis associated protein) (TAP) (Vesicle docking protein).  
 GN VDP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]

SEQUENCE FROM N.A., AND PHOSPHORYLATION SITE SER-942.  
 MEDLINE=98148093; PubMed=9478999;

RA Shohd M., Misumi Y., Yano A., Takami N., Ikehara Y.;  
 RT "Phosphorylation of the vesicle docking protein p115 regulates its association with the Golgi membrane."  
 RL J. Biol. Chem. 273:5385-5388(1998).

CC -1- FUNCTION: GENERAL VESICULAR TRANSPORT FACTOR REQUIRED FOR INTERCYSTERNAL TRANSPORT IN THE GOLGI STACK: IT IS REQUIRED FOR TRANSCYTIC FUSION AND/OR SUBSEQUENT BINDING OF THE VESICLES TO THE TARGET MEMBRANE. MAY WELL ACT AS A VESICULAR ANCHOR BY INTERACTING WITH THE TARGET MEMBRANE AND HOLDING THE VESICULAR AND TARGET MEMBRANES IN PROXIMITY (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN WHICH RECycles BETWEEN THE CYTOSOL AND THE GOLGI APPARATUS DURING INTERPHASE.  
 CC -1- DOMAIN: COMPOSED OF A GLOBULAR HEAD, AN ELONGATED TAIL (COILED-COIL) AND A HIGHLY ACIDIC C-TERMINAL DOMAIN.  
 CC -1- PTM: PHOSPHORYLATED IN A CELL CYCLE-SPECIFIC MANNER; PHOSPHORYLATED IN INTERPHASE BUT NOT IN MITOTIC CELLS.  
 CC DEPHOSPHORYLATED PROTEIN ASSOCIATES WITH THE GOLGI MEMBRANE; PHOSPHORYLATION PROMOTES DISSOCIATION.

CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: D86326; BAA25300.1; -  
 DR MIM: 603344; -  
 DR InterPro: IPR000225; Armadillo.  
 DR PROSITE: PS50176; ARM\_REPEAT; UNKNOWN\_1.  
 KW Transport; Protein transport; Golgi stack; Membrane; Coiled coil;  
 KW Phosphorylation.  
 FT DOMAIN 1 637 GLOBULAR HEAD.  
 FT DOMAIN 638 930 COILED COIL (POTENTIAL).  
 FT DOMAIN 935 962 ASP/GLU-RICH (ACIDIC).  
 FT MOD\_RES 942 942 ASP/PHOSPHORYLATION.  
 FT MUTAGEN 942 942 S->A: LOSS OF PHOSPHORYLATION.  
 SQ SEQUENCE 962 AA; 107906 MW; 2E748F2C1BCB2942 CRC64;

Query Match 11.1%; Score 99; DB 1; Length 962;  
 Best Local Similarity 28.5%; Pred. No. 1.8;

Matches 35; Conservative 26; Mismatches 54; Indels 8; Gaps 3;  
 QY 60 VMECRVYTHLLQOELTEAOKGFQDVE-AQAATCNHTVMALMASLDAEKAAGQKKVELEEG 118  
 Db 739 IEELKKNQELQSGLTEKDSMIMENKSSQTSNGNESSAIVSARDE-----QVALKQ 792  
 QY 119 EITTLNKKLQDASAEVERLRENO-VLSVRIADKKYPPSSODSSSAAPDQLIVLLGSA 177  
 Db 793 ELATLKSQLNSQVETIKQLTEKQELQTEAFAPKSVVEGETETIIVARTTVEGRISA 852  
 QY 178 LLQ 180  
 Db 853 LLQ 855

## RESULT 6

MYSB\_CAEEL STANDARD; PRT; 1966 AA.

ID MYSB\_CAEEL  
 AC P02366;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin heavy chain B (MHC B).  
 GN UNC-54 OR MYO-4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoidea;  
 OC Rhabdilitae; Peloderinae; Caenorhabdilitis.  
 OX NCBI\_TaxId=6239;  
 RN [1]

RA MEDLINE=83273600; PubMed=6576334;  
 RA Karn J., Brenner S., Barnett L.;  
 RT "Protein structural domains in the Caenorhabditis elegans unc-54 myosin heavy chain gene are not separated by introns."  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:4253-4257(1983).  
 RN [2]  
 RP SEQUENCE OF 850-1966 FROM N.A.  
 RX MEDLINE=82272395; PubMed=7202124;  
 RA McLachlan A.D., Karn J.;  
 RT "Periodic charge distributions in the myosin rod amino acid sequence RT match cross-bridge spacings in muscle."  
 RL Nature 299:226-231(1982).  
 RN [3]  
 RP SEQUENCE OF 1876-1966 FROM N.A.  
 RX MEDLINE=8332892; PubMed=6571695;  
 RA Wills N., Gesteland R.F., Karn J., Barnett L., Bolten S.,  
 RA Waterston R.H.;  
 RT "The genes sup-7 x and sup-5 III of C. elegans suppress amber nonsense mutations via altered transfer RNA."  
 RT Cell 33:575-583(1983).

CC -1- FUNCTION: MUSCLE CONTRACTION.  
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPEITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE SI DOMAIN ARE SELECTIVELY ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.

CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).  
 CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN C.ELEGANS.  
 CC -1- MISCELLANEOUS: MHC A AND MHC B ARE FOUND EXCLUSIVELY IN THE BODY WALL MUSCLE. THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.

CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: J01050; AAA28124.1; -  
DR EMBL: V01494; CAA24738.1; -  
DR PIR: A02992; MMKW.  
DR HSSP: P08799; 1MND.  
DR InterPro: IPR004009; Myosin\_N.  
DR InterPro: IPR002928; Myosin\_tail.  
DR InterPro: IPR001609; myosin\_head.  
DR Pfam: PF00063; myosin\_head.1.  
DR Pfam: PF02736; Myosin\_N.1.  
DR Pfam: PF01576; Myosin\_tail.1.  
DR PRINTS: PR00193; MYOSINHEAVY.  
DR ProDom: PD000355; myosin\_head.1.  
DR SMART: SM00242; MYSC.1.  
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;  
KW ATP-binding; Methylation; Alkylation; Multigene family.  
FT DOMAIN 1 850  
FT 1 850 COILED COIL (POTENTIAL).  
FT DOMAIN 851 1164 ALPHA-HELICAL TAILPIECE (S2).  
FT 851 1164 HINGE.  
FT DOMAIN 1165 1966 LIGHT MEROMYOSIN (LMM).  
FT NP\_BIND 177 184 ATP (BY SIMILARITY).  
FT DOMAIN 177 184 ACTIN-BINDING.  
FT DOMAIN 769 783 ACTIN-BINDING.  
FT MOD\_RES 128 128 METHYLATION (TRI-) (POTENTIAL).  
FT MOD\_RES 705 705 ALKYLATION (SH-1).  
FT MOD\_RES 715 715 ALKYLATION (SH-2).  
FT CONFLICT 1337 1337 E -> R (IN REF. 2).  
FT CONFLICT 1880 1880 I -> L (IN REF. 2).  
SQ SEQUENCE 1966 AA; 225125 MW; B66F0B2FE27B67F CRC64;

Query Match 11.1%; Score 98.5; DB 1; Length 1966;  
Best Local Similarity 24.7%; Pred. No. 4.2;  
Matches 36; Conservative 21; Mismatches 58; Indels 31; Gaps 5;

Oy 47 KANSEACGDLRAVMECRNVTLLDQELTEAGKGFQDYEAQAATCNHTVMALMASLDAEK 106  
Db 1811 EEAFAALKGKKVIAKLEQVRVLESELDGEQRFOQANKNGRADRVRELOFQVDEDK 1870  
Oy 107 AQ-----GCKVFELEGEITTLN-----HKLDASAEVERLRREN 141  
Db 1871 KNEERLQDLIDKLOQKLEKTKQKQVEAE-ELANLNLQKYKQLTQLEDAEERAD--QAEN 1927  
Oy 142 QVLAVRI--ADKRYPPSSQDSSAA 164  
Db 1928 SLSKMRSKRASASVAPGLQSSASAA 1953

RESULT 7  
ID MYH4\_RABIT STANDARD: PRT: 1938 AA.  
AC Q28641:  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin heavy chain, skeletal muscle, juvenile.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NEW ZEALAND WHITE; TISSUE=Skeletal muscle;  
RA Maeda K., Hostinova E., Roesch-Kleinlauf A., Schuster H., Gasperik J.,  
RA Wittinghofer A.;  
RT "Isolation, sequencing of myosin heavy chain cDNA from rabbit  
RT skeletal muscle and a novel cosynthesis of S-1 fragment with the  
RT essential and regulatory light chains.";

RL submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: MUSCLE CONTRACTION.  
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2  
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)  
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT  
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATTER BE  
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED  
CC SUBFRAGMENT (S2).  
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: U32574; AAA74199.1; -  
DR HSSP: P08799; 1MND.  
DR InterPro: IPR000048; IQ.  
DR InterPro: IPR004009; Myosin\_N.  
DR InterPro: IPR002928; Myosin\_tail.  
DR InterPro: IPR001609; myosin\_head.  
DR Pfam: PF00063; myosin\_head.1.  
DR Pfam: PF00612; IQ.2.  
DR Pfam: PF02736; Myosin\_N.1.  
DR Pfam: PF01576; Myosin\_tail.1.  
DR PRINTS: PR00193; MYOSINHEAVY.  
DR ProDom: PD000355; myosin\_head.1.  
DR SMART: SM00015; IQ.1.  
DR SMART: SM00242; MYSC.1.  
DR PROSITE: PS50096; IQ.1.  
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;  
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;  
KW Multigene family.  
FT DOMAIN 1 783  
FT 1 783 MYOSIN HEAD-LIKE.  
FT DOMAIN 784 813  
FT 784 813 IQ.  
FT NP\_BIND 842 1938 COILED COIL (POTENTIAL).  
FT 179 186 ATP (POTENTIAL).  
FT DOMAIN 658 680 ACTIN-BINDING (BY SIMILARITY).  
FT DOMAIN 760 774 ACTIN-BINDING (BY SIMILARITY).  
FT MOD\_RES 35 35 METHYLATION (MONO-) (BY SIMILARITY).  
FT MOD\_RES 130 130 METHYLATION (TRI-) (BY SIMILARITY).  
FT MOD\_RES 552 552 METHYLATION (TRI-) (BY SIMILARITY).  
FT MOD\_RES 756 756 METHYLATION (TRI-) (BY SIMILARITY).  
FT MOD\_RES 698 698 ALKYLATION (SH-1) (BY SIMILARITY).  
FT MOD\_RES 708 708 ALKYLATION (SH-2) (BY SIMILARITY).  
SQ SEQUENCE 1938 AA; 223064 MW; D8A8A2E5B182626 CRC64;

Query Match 11.0%; Score 98; DB 1; Length 1938;  
Best Local Similarity 26.1%; Pred. No. 4.6;  
Matches 37; Conservative 20; Mismatches 45; Indels 40; Gaps 5;

Oy 47 KANSEACR-----DGLRAVMECRNVTLLDQELTEAGKGFQDYEAQAATCNHTVMALM 99  
Db 1366 KANSEVAQWRIRYETDIAQTELEBAKKRIARLQDAE---EHVEAVAKC----- 1414  
Oy 100 ASLDAEKAQGGKVELEGEITTLN-----HKLDASAEVERLR 138  
Db 1415 ASLEKTKQKQLEVEDLMDIVERTNAACAALDKQKQNPDKIIAEKHKYEETHAELEASQ 1474  
Oy 139 RENQVLSRIAD-KRYPPSSQD 159  
Db 1475 KESRSISTEVFKVKNAYEESLD 1496

```

RESULT      8
MYSD_CAEEL  STANDARD; PRT; 1938 AA.
ID MYSD_CAEEL 019674;
AC P02567; 019674;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain D (MHC D).
GN MYO-1 OR R06C7.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=89178677; PubMed=2926820;
RZ Ddbb N.J., Maryama I.N., Krause M., Karn J.;
RT "Sequence analysis of the complete Caenorhabditis elegans myosin
   heavy chain gene family."
RL J. Mol. Biol. 205:603-613(1989).
RN [2]
RP SEQUENCE OF 34-1795 FROM N.A.
RX MEDLINE=83273600; PubMed=6576334;
RA Karn J., Brenner S., Barnett L.;
RT "Protein structural domains in the Caenorhabditis elegans unc-54
   myosin heavy chain gene are not separated by introns."
RL Proc. Natl. Acad. Sci. U.S.A. 80:4253-4257(1983).
RN [3]
RP SEQUENCE OF 115-365 AND 1492-1763 FROM N.A.
RX MEDLINE=85201409; PubMed=3888374;
RA Karn J., Ddbb N.J., Miller D.M.;
RT "Cloning nematode myosin genes."
RL Cell Muscle Motil. 6:185-237(1985).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX Gardner A., McMurtry A.;
RT Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RL
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE PHARYNGEAL MUSCLE.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN
CC C.ELEGANS.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X08065; CAA30854.1; -
CC EMBL: M37232; AAA28119.1; -
CC EMBL: M37234; AAA28120.1; -
CC EMBL: M71266; CAA95848.1; -
CC EMBL: M71261; CAA95848.1; JOINED.
CC EMBL: M71261; CAA95806.1; -

```

```

DR EMBL: M71266; CAA95806.1; JOINED.
DR PIR: S02772; MKKW1.
DR HSSP: P08799; 1MND.
DR WormPep: R06C7.10; CE06253.
DR InterPro: IPR004009; Myosin N.
DR InterPro: IPR002928; Myosin N.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head.
DR Pfam: PF02736; Myosin_N; 1.
DR Pfam: PF01576; Myosin_tail; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00242; MSc; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family.
FT DOMAIN 1 845
FT DOMAIN 846 1938
FT DOMAIN 846 1170
FT DOMAIN 1171 1938
FT NE_BIND 177 184
FT DOMAIN 660 682
FT DOMAIN 764 778
FT MOD_RES 128 128
FT MOD_RES 700 700
FT MOD_RES 710 710
FT MOD_RES 94 94
FT CONFLICT 98 98
FT CONFLICT 377 377
FT CONFLICT 389 390
FT CONFLICT 391 391
FT CONFLICT 408 408
FT CONFLICT 474 474
FT CONFLICT 577 577
FT CONFLICT 681 681
FT CONFLICT 1373 1373
FT CONFLICT 1659 1659
SQ SEQUENCE 1938 AA; 22325 MW; 387399C8F6344CF4 CRC64;
Query Match 11.0%; Score 98; DB 1; Length 1938;
Best Local Similarity 26.6%; Pred. No. 4.6;
Matches 34; Conservative 24; Mismatches 58; Indels 12; Gaps 4;
OY 51 EACRQDGLRAVME-----CRNYTHLQO--ELTEAQ-KGFQDVEAQAATCNHTVALMASLD 103
DB 1324 KAMEDLHENQERHACKNLEHEDQHELLEQINGKDIDQSLRINSEISQWARYE 1383
OY 104 AEKAGCKKVEELGELTTLNHKLQDASAVEKLRRENOYLSVRIADKTYPSQDSSSA 163
DB 1384 GEGLYGSEELDELRKQMNREVMDOEALSA-----QNKVLSLEKAKGKLLETEDARSD 1438
OY 164 AAPQLLIV 171
DB 1439 VDRHLTYI 1446
RESULT      9
P092_SCICO  STANDARD; PRT; 286 AA.
ID P092_SCICO 022312;
AC P22312;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE PUFF 11/9-2 protein precursor.
GN PUFF 11/9-2.
OS Sciatara coprophila (Fungus gnat).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Sciarioidea;
OC Sciaridae; Bradysia.
OX NCBI_Taxid=38358;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN-6980;  
 RX MEDLINE=90133907; PubMed=2614832;  
 RA Dibratolomeis S.M., Gerbi S.A.;  
 RT "Molecular characterization of DNA puff II/9A genes in Sclara  
 copophila.";  
 RL J. Mol. Biol. 210:531-540(1989).  
 CC -I- MISCELLANEOUS: THE DNA PUFF II/9 PROTEINS HAVE A PROPOSED  
 CC INTERMOLECULAR COILED COIL STRUCTURE WITH POSSIBLY INTERMOLECULAR  
 CC DISULFIDE BRIDGES FORMED BY NUMEROUS CYSTEINE RESIDUES IN POSITION  
 CC D OF THE HEPTAD REPEAT.  
 CC -I- SIMILARITY: 76% IDENTICAL TO THE PUFF II/9-1 PROTEIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X51679; CAA35982.1; -  
 CC PIR: S07533; S07533.  
 CC Signal; coiled coil; glycoprotein.  
 CC SIGNAL 1 19 OR 21 (POTENTIAL).  
 CC CHAIN 1 20 PUFF II/9-2 PROTEIN.  
 CC DOMAIN 61 235 HELICAL (POTENTIAL).  
 CC CAROHD 156 156 N-LINKED (GLCNAC - ) (POTENTIAL).  
 CC SEQUENCE 286 AA; 32621 MW; 720AC8CC22A869C CRC64;

Query Match 10.9%; Score 97; DB 1; Length 286;  
 Best Local Similarity 25.6%; Pred. No. 0.7;  
 Matches 31; Conservative 24; Mismatches 36; Indels 30; Gaps 5;

OY 46 IANSEACDGLRAVMEC-RNVTHL-----IQDELTEAKGFQDVEAQAATCNHTMAL 98  
 DB 89 IREKREARQAKAEKALKECKNTENLKEETEQLKEAEKQALKECKKELADCK----- 142  
 OY 99 MASLAEKQAGKQVELEGETTLNHRQ-----DASAVERLRRE-----NOVLSV 146  
 DB 143 -----KENKLLNKTEELNCTTLOELKRCGRERDLOCLDECKKLLNCTNELLAC 197  
 OY 147 R 147  
 DB 198 R 198

RESULT 10

VDP\_RAT  
 VDP\_RAT STANDARD: PRT; 959 AA.  
 P41542;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE General vesicular transport factor p115 (Transcytosis associated  
 DE protein) (TAP) (Vesicle docking protein).  
 GN VDP  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=95132632; PubMed=7831323;  
 RA Saperstein S.K., Walter D.M., Grosvenor A.R., Heuser J.E.,  
 RA Waters M.G.;  
 RT "p115 is a general vesicular transport factor related to the yeast  
 RT endoplasmic reticulum to Golgi transport factor Usolp.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:522-526(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Liver;

RX MEDLINE=95132633; PubMed=7831324;  
 RA Barroso M., Nelson D.S., Szul E.;  
 RT "Transcytosis-associated protein (TAP)/p115 is a general fusion  
 RT factor required for binding of vesicles to acceptor membranes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:527-531(1995).  
 CC -I- FUNCTION: GENERAL VESICULAR TRANSPORT FACTOR REQUIRED FOR  
 CC INTERCISTERNAL TRANSPORT IN THE GOLGI STACK; IT IS REQUIRED FOR  
 CC TRANSCYTIC FUSION AND/OR SUBSEQUENT BINDING OF THE VESICLES TO  
 CC THE TARGET MEMBRANE. MAY WELL ACT AS A VESICULAR ANCHOR BY  
 CC INTERACTING WITH THE TARGET MEMBRANE AND HOLDING THE VESICULAR  
 CC AND TARGET MEMBRANES IN PROXIMITY.  
 CC -I- SUBUNIT: HOMODIMER WITH TWO N-TERMINAL HEADS AND A C-TERMINAL  
 CC COILED-COIL TAIL. DIMER FORMED BY PARALLEL ASSOCIATION OF THE  
 CC TAILS.  
 CC -I- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN WHICH RECYCLES  
 CC BETWEEN THE CYTOSOL AND THE GOLGI APPARATUS DURING INTERPHASE.  
 CC -I- DOMAIN: COMPOSED OF A GLOBULAR HEAD, AN ELONGATED TAIL (COILED-  
 CC COIL) AND A HIGHLY ACIDIC C-TERMINAL DOMAIN.  
 CC -I- PTM: PHOSPHORYLATED IN A CELL CYCLE-SPECIFIC MANNER;  
 CC PHOSPHORYLATED IN INTERPHASE BUT NOT IN MITOTIC CELLS.  
 CC DEPHOSPHORYLATED PROTEIN ASSOCIATES WITH THE GOLGI MEMBRANE;  
 CC PHOSPHORYLATED PROMOTES DISSOCIATION (BY SIMILARITY).  
 CC -I- SIMILARITY: BELONGS TO THE VDP/USOL/YB1047C FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U14192; AA62632.1; -  
 CC EMBL: U15389; AAC52151.1; -  
 CC InterPro: IPR000225; Armadillo.  
 CC DR PROSITE: PS50176; ARM\_REPEAT; UNKNOWN\_1.  
 CC TRANSPORT: Protein transport; Golgi stack; Membrane; Coiled coil;  
 CC phosphorylation.  
 CC DOMAIN 1 637 GLOBULAR HEAD.  
 CC FT DOMAIN 638 930 COILED COIL (POTENTIAL).  
 CC FT DOMAIN 935 959 ASP/GLU-RICH (ACIDIC).  
 CC FT MOD\_RES 940 940 PHOSPHORYLATION (BY SIMILARITY).  
 CC FT CONFLICT 591 591 S -> P (IN REF. 2).  
 CC FT CONFLICT 658 658 M -> V (IN REF. 2).  
 CC FT CONFLICT 816 816 S -> R (IN REF. 2).  
 CC FT CONFLICT 873 873 A -> S (IN REF. 2).  
 CC SEQUENCE 959 AA; 107162 MW; 356394B48C7E03B CRC64;

Query Match 10.9%; Score 97; DB 1; Length 959;  
 Best Local Similarity 30.1%; Pred. No. 2.6;  
 Matches 37; Conservative 20; Mismatches 58; Indels 8; Gaps 3;

OY 60 VNECRVNTLLLOELTEAKGFQDV-EAQAATCNHTMALMSLAEKQAGKQVELEBG 118  
 DB 739 IEIRSHOVLDSQALAEKQVTEENLRSSQVSGMSEBALTCSPRDE-----QVAELKQ 792  
 OY 119 EITTLNHLQDASAVERLRRENOVLSVRAD-KKYPSSQDSSSAADQLITVLGISA 177  
 DB 793 ELSALKSQCSQSLFETTRQTEENSELQRAETLAKSPVEGSELVTAKTADVEGRISA 852

OY 178 LQ 180  
 DB 853 LQ 855

RESULT 11  
 YM92\_CAEFL  
 ID YM92\_CAEFL STANDARD: PRT; 893 AA.  
 AC P34531; P34532; P34533;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)

```

DE Hypothetical 100.0 kDa protein M01A8.2 in chromosome III.
GN M01A8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Spratt J.,
RA Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z27081; CAA81607.1; -.
DR PIR: S40998; S40998.
DR PIR: S40999; S40999.
DR WormPep: M01A8.2; CE03491.
DR InterPro: IPR000938; CAP-GLY.
DR Pfam: PF01302; CAP-GLY.1.
DR PROSITE: PS00845; CAP-GLY.1; 1.
DR PROSITE: PS50245; CAP-GLY.2; 1.
DR Hypothetical protein; Coiled coil.
FT DOMAIN 39 81 CAP-GLY.
FT DOMAIN 522 696 COILED COIL (POTENTIAL).
FT DOMAIN 729 756 COILED COIL (POTENTIAL).
FT SEQUENCE 893 AA; 99997 MW; 464F2962B36C28B1 CRC64;

Query Match 10.9%; Score 96.5; DB 1; Length 893;
Best Local Similarity 24.8%; Pred. No. 2.6;
Matches 32; Conservative 29; Mismatches 39; Indels 29; Gaps 4;

QY 48 ANSACRDGLRAVMECRNVTLLQ--QELTEAKGFQDVA--QAQATCNHTYMALMAS 101
DB 566 SNOOVYNNHNAVH-ESLQKTHEIOLIAKKNKEFEERFEERARRAEVCAMNRQKVVAC 624
QY 102 LDAEKAQOKKVEELE-----GETITLNHKIADASAEVERLR 139
DB 625 LDEISIAEAKQCEQLNDKVKYLAALANDCDHRNMLTKETISLOTALAEKSAEMKELRQ 684
QY 140 ENQVLSVRI 148
DB 685 KNQNLISLQV 693

RESULT 12
P091_SCICO STANDARD; PRT; 286 AA.
AC P22311;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE PUFF II/9-1 protein precursor.
GN II/9-1.
OS Sciara coprophila (Fungus gnaf).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pelegrina; Neoptera; Endopterygota; Diptera; Nematocera; Sciarioidea;
OC Sciariidae; Brachystia.
OX NCBI_TaxId=38358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6980;
RX MEDLINE=90133907; PubMed=2614832;
RA Dibatolomels S.M., Gerbi S.A.;
RA "Molecular characterization of DNA puff II/9A genes in Sciara
RA coprophila.";
RT J. Mol. Biol. 210:531-540(1989).
CC -1- MISCELLANEOUS: THE DNA PUFF II/9 PROTEINS HAVE A PROPOSED
CC INTERMOLECULAR COILED COIL STRUCTURE WITH POSSIBLY INTERMOLECULAR
CC DISULFIDE BRIDGES FORMED BY NUMEROUS CYSTEINE RESIDUES IN POSITION
CC D OF THE HEPTAD REPEAT.
CC -1- SIMILARITY: 76% IDENTICAL TO THE PUFF II/9-2 PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X51680; CAA35983.1; -.
DR PIR: S07532; S07532.
DR Signal; Coiled coil; Glycoprotein.
FT SIGNAL 1 19 OR 21 (POTENTIAL).
FT CHAIN 20 286 PUFF II/9-1 PROTEIN.
FT DOMAIN 61 235 HELICAL (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC... ) (POTENTIAL).
FT SEQUENCE 286 AA; 32034 MW; AA6A7B5F5191BBID CRC64;

Query Match 10.8%; Score 96; DB 1; Length 286;
Best Local Similarity 30.0%; Pred. No. 0.84;
Matches 24; Conservative 16; Mismatches 36; Indels 4; Gaps 1;

QY 59 AVMECRNVTLLQDELTEAKGFQDVAQAATCNHTYMALMASLDAEKAQOKKVEELG 118
DB 102 ALCECQNSSELKQTIQLKKELAQTQELANCKEA----LANCKAENAKLTKIEELNG 157
QY 119 EITTLNKHKIDASAEVERLR 138
DB 158 TITQLELEQCRARERDQ 177

RESULT 13
MYH8_HUMAN STANDARD; PRT; 1937 AA.
AC P13535; Q14910;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, skeletal muscle, perinatal (MyHC-perinatal).
GN MYH8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=90323631; PubMed=2373371;
RA Karsch-Mizrachi I., Peghali R., Shows T.B., Jr., Leinwand L.A.;
RA "Generation of a full-length human perinatal myosin heavy-chain-
RA encoding cDNA.";
RT Gene 89:289-294(1990).

```

[2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=95324556; PubMed=7601129;  
 RA Julian E.H., Kelly A.M., Pompidou A.J., Hoffman R., Schiaffino S.,  
 RT Stehman H.H., Rudenstein N.A.;  
 RT "Characterization of a human perinatal myosin heavy-chain  
 transcript.";  
 RL Eur. J. Biochem. 230:1001-1006(1995).  
 [3]  
 RP SEQUENCE OF 502-1937 FROM N.A.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=90235862; PubMed=1691980;  
 RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goede H.W.,  
 RT Arnold H.H.;  
 RT "Identification of three developmentally controlled isoforms of human  
 myosin heavy chains.";  
 RL Eur. J. Biochem. 189:55-65(1990).  
 [4]  
 RP SEQUENCE OF 860-1937 FROM N.A.  
 RX MEDLINE=89234168; PubMed=2715179;  
 RA Peghali R., Leinwand L.A.;  
 RT "Molecular genetic characterization of a developmentally regulated  
 human perinatal myosin heavy chain.";  
 RL J. Cell Biol. 108:1791-1797(1989).  
 [5]  
 RP SEQUENCE OF 1-46 FROM N.A.  
 RA Esser K., Tidhar A., Myszkowski M.;  
 RT "Isolation and characterization of the human perinatal MHC promoter.";  
 RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: MUSCLE CONTRACTION.  
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2  
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)  
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY  
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.  
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT  
 MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE  
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED  
 SUBFRAGMENT (S2).  
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M36769; AAC17185.1; -;  
 DR EMBL: 238133; CAAB6293.1; -;  
 DR EMBL: X51593; CAAB3594.1; -;  
 DR EMBL: M35250; AAA36346.1; -;  
 DR EMBL: AF067143; AAC21557.1; -;  
 DR PIR: A30220; A30220.  
 DR HSSP: P08799; A30220.  
 DR MIM: 160741; -;  
 DR InterPro: IPR000048; IO.  
 DR InterPro: IPR004009; Myosin\_N.  
 DR InterPro: IPR002928; Myosin\_tail.  
 DR InterPro: IPR001609; myosin\_head.  
 DR Pfam: PF00612; IO; 1.  
 DR Pfam: PF00063; myosin\_head; 1.  
 DR Pfam: PF02736; Myosin\_N; 1.  
 DR Pfam: PF01576; Myosin\_tail; 1.  
 DR PRINTS: PR00193; MYOSINHEAVY.  
 DR ProDom: PD000355; myosin\_head; 1.

DR SMART; SM00015; IO; 1.  
 DR SMART; SM00242; MYSC; 1.  
 DR PROSITE; PS50096; IO; 1.  
 KW Myosin: Muscle protein, coiled coil; Thick filament; Actin-binding;  
 KW ATP-binding; Methylation; Alkylation; Multigene family;  
 KW Calmodulin-binding.  
 FT DOMAIN 1 780  
 FT MYOSIN HEAD-LIKE.  
 FT IO.  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 842 1937  
 FT ATP.  
 FT NP\_BIND 181 188  
 FT ACTIN-BINDING.  
 FT DOMAIN 658 680  
 FT ACTIN-BINDING.  
 FT MOD\_RES 760 774  
 FT METHYLATION (TR1) (POTENTIAL).  
 FT MOD\_RES 132 132  
 FT ALKYLATION (SH-1) (POTENTIAL).  
 FT MOD\_RES 698 698  
 FT ALKYLATION (SH-2) (POTENTIAL).  
 FT MOD\_RES 708 708  
 FT A -> R (IN REF. 2).  
 FT CONFLICT 15 15  
 FT E -> Q (IN REF. 1 AND 4).  
 FT CONFLICT 970 970  
 FT M -> N (IN REF. 1 AND 4).  
 FT CONFLICT 1072 1072  
 FT N -> H (IN REF. 3).  
 FT CONFLICT 1247 1247  
 FT MC -> DG (IN REF. 3).  
 FT CONFLICT 1251 1252  
 FT E -> G (IN REF. 1 AND 4).  
 FT CONFLICT 1261 1261  
 FT K -> Q (IN REF. 1 AND 4).  
 FT CONFLICT 1297 1297  
 FT K -> NT (IN REF. 3).  
 FT CONFLICT 1378 1378  
 FT EN -> AH (IN REF. 1 AND 4).  
 FT CONFLICT 1504 1505  
 FT E -> D (IN REF. 1 AND 4).  
 FT CONFLICT 1847 1847  
 FT D -> H (IN REF. 2).  
 FT CONFLICT 1914 1914  
 SO SEQUENCE 1937 AA; 222762 MW; A3EE2D151792E9E8 CRC64;  
 Query Match 10.7%; Score 95; DB 1; Length 1937;  
 Best Local Similarity 25.2%; Pred. No. 7.9;  
 Matches 33; Conservative 21; Mismatches 59; Indels 18; Gaps 4;  
 QY 47 KANSEACR-----DGLRAVMECRNVTHLLDOELTEAQKGVDEAQAATCNHTVMALM 99  
 DB 1366 KANSEVAQMRKTYEDPAIORTTELEAKKKRLQRLQEAHEHVAVNAKASLEKTKORIQ 1425  
 QY 100 -----ASLAAKNAQ-----OKKVELEGEETITLNNKLDASAEVRLRENOVLSVRIA 149  
 DB 1426 NEVEDIMLVERSNAAALDKKRNFDVLSBWKOKYEETOAEHASKESRSJSTELP 1485  
 QY 150 D-KKYPSSOD 159  
 DB 1486 KVKKNYEESLD 1496  
 RESULT 14  
 MYHB\_HUMAN STANDARD; PRT; 1972 AA.  
 AC P35749; 000396; P78422; O94944;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin heavy chain, smooth muscle isoform (SMMHC).  
 GN MYH1 OR KIAA0866.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94425270; PubMed=10493829;  
 RA Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,  
 RA Fuhrman J., Mason T., Crosby M.L., Barnstead M., Cronin L.,  
 RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,  
 RA Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;  
 RT "Genome duplications and other features in 12 Mb of DNA sequence from  
 human chromosome 16p and 16q.";  
 RL Genomics 60:295-308(1999).  
 [2]  
 RP SEQUENCE OF 1-1266 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99156230; PubMed=10048485;

RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XII.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 5:355-364(1998).  
RN [3]  
RP SEQUENCE OF 885-1972 FROM N.A.  
RX MEDLINE=93263189; PubMed=7684189;  
RA Matsuda R., Yoshida M.C., Furutani Y., Imamura S., Kanda N.,  
RA Yanagisawa M., Masaki T., Takao A.;  
RT "human smooth muscle myosin heavy chain gene mapped to chromosomal  
RT region 16q12.";  
RL Am. J. Med. Genet. 46:61-67(1993).  
RN [4]  
RP SEQUENCE OF 1093-1972 FROM N.A.  
RC TISSUE=Hippocampus;  
RA Okajima K.;  
CC Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: MUSCLE CONTRACTION.  
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2  
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)  
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE; EXPRESSED IN THE UMBILICAL  
CC ARTERY, BLADDER, ESOPHAGUS AND TRACHEA.  
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY  
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.  
CC -1- DISEASE: A CHROMOSOMAL REARRANGEMENT, KNOWN AS PERICENTRIC  
CC INVERSION INV(16)(P13Q22), PRODUCES A FUSION PROTEIN THAT CONSISTS  
CC OF THE 165 N-TERMINAL RESIDUES OF CBF-BETA (PEP3) WITH THE TAIL  
CC REGION OF MYH11. THIS REARRANGEMENT IS ASSOCIATED WITH ACUTE  
CC MYELOID LEUKEMIA OF MAEO SURTYPE.  
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT  
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE  
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED  
CC SUBFRAGMENT (S2).  
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF001548; AAC31665.1; -;  
DR EMBL: U91323; AAC35212.1; -;  
DR EMBL: AB020673; BAA74889.1; -;  
DR EMBL: D10667; -. NOT\_ANNOTATED\_CDS.  
DR EMBL: X69292; CAA49134.1; -;  
DR HSP: P08799; IMMN.  
DR MIM: 160745; -;  
DR InterPro: IPR000048; IQ.  
DR InterPro: IPR002928; Myosin\_tail.  
DR InterPro: IPR002017; Spectrin.  
DR InterPro: IPR001609; myosin\_head.  
DR Pfam: PF00612; IQ; 1.  
DR Pfam: PF00063; myosin\_head; 1.  
DR Pfam: PF01576; myosin\_tail; 1.  
DR PRINTS: PR00193; MYOSINHEAVY.  
DR PRODOM: PD000355; myosin\_head; 1.  
DR SMART: SM00015; IQ; 2.  
DR SMART: SM00242; MISC; 1.  
DR PROSITE: PS50096; IQ; 1.  
DR Myosin: Muscle protein: Coiled coil; Thick filament; Actin-binding;  
KM Calmodulin-binding; ATP-binding; Methylation; Alkylation;  
KM Multigene family; Proto-oncogene; Chromosomal translocation.

FT DOMAIN 1 785 MYOSIN HEAD-LIKE.  
FT DOMAIN 786 815 IQ.  
FT DOMAIN 844 1934 COILED COIL (POTENTIAL).  
FT DOMAIN 1935 1972 CARBOXYL-TERMINAL.  
FT NP\_BIND 178 185 ATP (POTENTIAL).  
FT DOMAIN 661 683 ACTIN-BINDING (BY SIMILARITY).  
FT DOMAIN 762 776 ACTIN-BINDING (BY SIMILARITY).  
FT MOD\_RES 129 129 METHYLATION (TR1-) (POTENTIAL).  
FT MOD\_RES 701 701 ALKYLATION (SH-1) (POTENTIAL).  
FT MOD\_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).  
FT CONFLICT 887 889 EEK -> NSE (IN REF. 3).  
FT CONFLICT 1263 1266 ELOS -> TTSF (IN REF. 2).  
FT CONFLICT 1558 1558 T -> S (IN REF. 3).  
FT CONFLICT 1610 1611 KQ -> NE (IN REF. 3).  
FT CONFLICT 1786 1786 A -> S (IN REF. 4).  
FT CONFLICT 1958 1958 T -> L (IN REF. 3).  
SQ SEQUENCE 1972 AA; 227338 MW; 676658B2ACE1277 CRC64;  
  
Query Match 10.7%; Score 95; DB 1; Length 1972;  
Best Local Similarity 24.5%; Pred. No. 8;  
Matches 27; Conservative 18; Mismatches 55; Indels 10; Gaps 1;  
  
QY 53 CROSLAVMECRNVTHTLLOELTEAOKGFQDVEAQAATCNHTVAALASID----- 103  
DB 1268 CSDEERARAEINDKVHKLONEVESVTGMLEAEKAKIKAKDVASLSQDQDELQEE 1327  
QY 104 -AEKAGQKVEELEGETITLHNKLQDASAEVERRENOVLSVRIADK 152  
DB 1328 TRQKLVSTKLRLQLEERNSLDQDLDEMEKAKONLEHISTINQLSDSK 1377  
  
RESULT 15  
MYHB\_RABIT  
ID MYHB\_RABIT STANDARD; PRT; 1972 AA.  
AC P35748;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin heavy chain, smooth muscle isoform (SMHMC).  
GN MYH11.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
ON NCBI\_TaxID=9986;  
RX MEDLINE=92073350; PubMed=1961735;  
RA Babić P., Kelly C., Perissamy M.;  
RT "Characterization of a mammalian smooth muscle myosin heavy-chain  
RT gene: complete nucleotide and protein coding sequence and analysis of  
RT the 5' end of the gene.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10676-10680(1991).  
CC -1- FUNCTION: MUSCLE CONTRACTION.  
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2  
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)  
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY  
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.  
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT  
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE  
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED  
CC SUBFRAGMENT (S2).  
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its





**THIS PAGE BLANK (USPTO)**

Result No.	Score	Query Match	Length	DB	ID	Description
1	889	100.0	180	2	A56836	bone marrow stromal
2	111	12.5	778	2	T30430	hypothetical prote
3	108	12.1	421	2	T59463	keratin, type I, c
4	107.5	12.1	463	2	H84069	cell wall-binding
5	107	12.0	143	2	G83955	flagellar protein
6	102.5	11.5	1534	2	A56734	ribosome receptor
7	102	11.5	550	2	A28166	Kupffer cell recep
8	99.5	11.2	304	2	JX0209	lectin, galactose
9	99	11.1	415	2	S35760	fcra protein precu
10	98.5	11.1	1965	1	MMKN	myosin heavy chain
11	98	11.0	959	2	A53913	transcytosis-assoc
12	98	11.0	1938	1	MMKN	myosin heavy chain
13	98	11.0	1938	2	A59293	skeletal myosin he
14	97	10.9	286	2	S07533	puff II/9A-2 prote
15	96.5	10.9	597	2	S40998	hypothetical prote
16	96.5	10.9	893	2	G88551	protein M01A8.2 (H
17	96	10.8	286	2	S07532	puff II/9-1 protei
18	95.5	10.7	173	2	S76705	hypothetical prote
19	95	10.7	1937	2	T38055	myosin heavy chain
20	95	10.7	1972	1	A41604	myosin heavy chain
21	94.5	10.6	1388	2	S74245	serine/threonine-s
22	94	10.6	676	2	S00084	myosin heavy chain
23	93.5	10.5	140	2	H64629	hypothetical prote
24	93.5	10.5	1938	1	S06005	myosin alpha heavy
25	93.5	10.5	1939	2	T48175	myosin heavy chain
26	93	10.5	359	2	T50712	TOP AP - chicken
27	93	10.5	848	2	A44972	parmyosin - nema
28	93	10.5	879	2	A48573	parmyosin - nema
29	93	10.5	1938	2	JC5421	smooth muscle myos

30	93	10.5	1972	2	JC5420	smooth muscle myosin
31	92.5	10.4	392	2	G95258	secreted 45 kd pro
32	92.5	10.4	382	2	B98124	general stress protein
33	92.5	10.4	1509	1	A27224	myosin heavy chain
34	92.5	10.4	1598	1	A49464	alpha cardiac myosin
35	92	10.3	746	2	T47237	myosin II heavy chain
36	92	10.3	866	2	S04027	paramyosin - Caenorhabditis
37	92	10.3	872	2	T19296	hypothetical protein
38	91.5	10.3	244	2	I36913	beta-myosin heavy chain
39	91.5	10.3	764	2	I51302	myosin heavy chain
40	91.5	10.3	1039	2	S18199	myosin heavy chain
41	91.5	10.3	1934	2	I48153	myosin heavy chain
42	91.5	10.3	1935	1	A37102	myosin beta heavy chain
43	91.5	10.3	1935	1	S06006	myosin beta heavy chain
44	91.5	10.3	1939	1	A46762	myosin alpha heavy chain
45	91.5	10.3	2442	2	T08621	centrosome associated

## ALIGNMENTS

```

RESULT      1      '
A:56836
bone marrow stromal cell surface protein BST-2 - human
C:Species: Homo sapiens (man)
C:Date: 11-Aug-1995 #sequence.revision 11-Aug-1995 #text.change 21-Jul-2000
C:Accession: A56836
R:Rislikwa, J.; Kishino, T.; Tomizawa, H.; Lee, B.O.; Kobune, Y.; Inazawa, J.; Oritani
Genomics 26, 527-534, 1995
A:Title: Molecular cloning and chromosomal mapping of a bone marrow stromal cell surf
A:Reference number: A56836; MUID:95331788
A:Accession: A56836
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-180 <ISH>
A:Cross-references: GB:D28137; NID:g457563; PIDN:BAA0567.1; PID:g506861
C:Genetics:
A:Gene: GDB:BST2
A:Cross-references: GDB:409946; OMIM:600534
A:Map position: 19p13.2-19p13.2
C:Keywords: transmembrane protein

```

Query Match	100.0%;	Score 889;	DB 2;	Length 180;
Best Local Similarity	100.0%;	Pred. No. 2.6e-62;		
Matches 180;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

  

QY	1	MASTSYDYCRYPMEDGDKRCKLLLGIGILVLLIIVILGVPPLIIIFTIKANSEACRDLGLAV	60
Db	1	MASTSYDYCRYPMEDGDKRCKLLLGIGILVLLIIVILGVPPLIIIFTIKANSEACRDLGLAV	60
QY	61	MECRNVTHLLOQLTEAOKGFQDVEAQAATCNHTYMALMASIDAERKAQOKVVELEGEI	120
Db	61	MECRNVTHLLOQLTEAOKGFQDVEAQAATCNHTYMALMASIDAERKAQOKVVELEGEI	120
QY	121	TTLNHKLDDASAEVRLRENOVLSVRITADKKYTPSSODSSSAAPOLLIVILGSLALLQ	180
Db	121	TTLNHKLDDASAEVRLRENOVLSVRITADKKYTPSSODSSSAAPOLLIVILGSLALLQ	180

  

RESULT	2
T30430	
hypothetical protein ORF82 - Lymantria dispar nuclear polyhedrosis virus	
C:Species: Lymantria dispar nuclear polyhedrosis virus, LDMNPV	
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999	
C:Accession: T30430	
R:Virolo, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slaveick, J.M.; F	
A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria	
A:Reference number: Z20836; MUID:99124785	
A:Accession: T30430	
A:Status: preliminary; translated from GB/EMBL/DBD	
A:Molecule type: DNA	



QY 75 TEA-----OKGFQDVEAQAATCNHTVAMLSIDAEGAOKKVELEGTEITTLNKK 126  
 Db 66 VRARLHVDRNMLAQEVQKRFQYNNITITSLAGMDASQSLSGAGICKLEAQ----- 118  
 QY 127 LODASAEVERLRRENOVL 146  
 Db 119 LOD--DINALMOEGSVYKI 135

RESULT 6  
 A56734  
 ribosome receptor, 180k - dog  
 C:Species: Canis lupus familiaris (dog)  
 C>Date: 08-Jul-1995 #sequence\_revision 03-Aug-1995 #text\_change 08-Oct-1999  
 C:Accession: A56734  
 R:Maner, E.E., Sun, Y., Savitz, A.J., Meyer, D.I.,  
 J. Cell Biol. 130, 29-39, 1995  
 A:Title: Functional characterization of the 180-kD ribosome receptor in vivo.  
 A:Reference number: A56734; MUID:95310363  
 A:Accession: A56734  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-1534 <MAN>  
 A:Cross-references: GB:X87224; NID:g984113; PIDN:CAA60676.1; PID:g984114  
 C:Keywords: endoplasmic reticulum; membrane protein; protein biosynthesis  
 F:198-743/Region: 10-residue repeats (N-Q-G-K-K-A-E-G-A-P)

Query Match 11.5%; Score 102.5; DB 2; Length 1534;  
 Best Local Similarity 28.6%; Pred. No. 4.4;  
 Matches 26; Conservative 22; Mismatches 42; Indels 1; Gaps 1;  
 QY 55 DGLRAVMGCRNVTHTLLOEFLTEAOKGFQDVEAQAATCNHTVAMLSIDAEGAOKKVE 114  
 Db 847 DAAVAKSKLREVNKELAEKAAAGAKVKQVLVAREOETTVAVARIEASYREHVKEVO 906  
 QY 115 ELEGEITTLNKKLQDA-SAEVERLRRENOVL 144  
 Db 907 QLOGKIRTLQEGLENGPNTQLARLQGENSIL 937

RESULT 7  
 A28166  
 Kupffer cell receptor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 05-Nov-1999  
 C:Accession: A28166; A28166  
 R:Hoyle, G.W., Hill, R.L.  
 J. Biol. Chem. 266, 1850-1857, 1991  
 A:Title: Structure of the gene for a carbohydrate-binding receptor unique to rat Kupffer  
 A:Reference number: A28166; MUID:91107689  
 A:Accession: A28166  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-550 <HO2>  
 A:Cross-references: GB:M55532; NID:g203362; PIDN:AAA40892.1; PID:g203363  
 R:Hoyle, G.W., Hill, R.L.  
 J. Biol. Chem. 263, 7487-7492, 1988  
 A:Title: Molecular cloning and sequencing of a cDNA for a carbohydrate binding receptor  
 A:Reference number: A28166; MUID:88227939  
 A:Accession: A28166  
 A:Molecule type: mRNA  
 A:Residues: 1-550 <HOV>  
 A:Cross-references: GB:J03734; NID:g205050; PIDN:AAA41472.1; PID:g205051  
 C:Superfamily: C-type lectin homology  
 C:Keywords: transmembrane protein  
 F:412-536/Domain: C-type lectin homology <LCH>

Query Match 11.5%; Score 102; DB 2; Length 550;  
 Best Local Similarity 28.0%; Pred. No. 1.7;  
 Matches 37; Conservative 20; Mismatches 61; Indels 14; Gaps 3;

QY 46 IKANSEACRDIIRAVMECRNVTHTLLOEFLTEAOKGFQDVEAQAATCNHTVAMLSIDA 105  
 Db 256 ISAEIQARBDGMGRAGE-----EMTSIKDLEFLTQIQNANHILQDTQIQGL 305  
 QY 106 KAQOKKVELEGTEITTLNKKLQDASAEVERLRRENOVL 146  
 Db 306 KAQ-LKSTSSINSQIEVNVGKTKDSSRELQTLRRD---LSDVSAKSNVQMQLQAK 361  
 QY 166 POLLIVILGLSA 177  
 Db 362 AEVOSLKTGLEA 373

RESULT 8  
 JX0209  
 lectin, galactose/N-acetylgalactosamine-specific - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 20-Aug-1999  
 C:Accession: JX0209; PX0009  
 R:Sato, M.; Kawakami, K.; Osawa, T.; Toyoshima, S.  
 J. Biochem. 111, 331-336, 1992  
 A:Title: Molecular cloning and expression of cDNA encoding a galactose/N-acetylgalact  
 A:Reference number: JX0209; MUID:92268032  
 A:Accession: JX0209  
 A:Molecule type: mRNA  
 A:Residues: 1-304 <SAT>  
 A:Cross-references: GB:S36676; NID:g249360; PIDN:AA22171.1; PID:g249361  
 R:Oda, S.; Sato, M.; Toyoshima, S.; Osawa, T.  
 J. Biochem. 104, 600-605, 1988  
 A:Title: Purification and characterization of a lectin-like molecule specific for gal  
 A:Reference number: PX0009; MUID:89197865  
 A:Accession: PX0009  
 A:Molecule type: protein  
 A:Residues: 102-120;137,X',139-151 <ODA>  
 C:Superfamily: hepatic lectin; C-type lectin homology  
 C:Keywords: glycoprotein; lectin; macrophage; transmembrane protein  
 F:36-61/Domain: transmembrane #status predicted <TRA>  
 F:173-296/Domain: C-type lectin homology <LCH>  
 F:74,166/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 11.2%; Score 99.5; DB 2; Length 304;  
 Best Local Similarity 23.2%; Pred. No. 1.4;  
 Matches 36; Conservative 37; Mismatches 49; Indels 33; Gaps 7;  
 QY 24 LGIGLIVLITVLCVPLIIFTIKANSEACRD--GLRAVMGCRNVTHTLLOEFLTEAOKGF 81  
 Db 41 LGLSLLLVAVSVIG-----SONQLRDLGTLRATLD--NTTSRIKAE-----F 83  
 QY 82 ODVEAQAATCNHTVAMLSIDAEGAOKG-----KVELEGTEITTLNKKLQDAS 131  
 Db 84 OSLSRRADSPFEKGISLKYVDHROELQAGDLSQKYSLSSTVEKREBQALKTPLSLTL 143  
 QY 132 AEVERLRRENOVL 146  
 Db 144 DHVQQLRDLKALTLQGLANLK-----NMSSEVACCP 174

RESULT 9  
 S35760  
 forA protein precursor - Streptococcus pyogenes  
 C:Species: Streptococcus pyogenes  
 C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 26-Aug-1999  
 C:Accession: S35760; A42711  
 R:Podbielski, A.  
 submitted to the EMBL Data Library, November 1992  
 A:Reference number: S35760  
 A:Accession: S35760  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-415 <POD>  
 A:Cross-references: EMBL:X69324; NID:g311759; PIDN:CAA49165.1; PID:g311760  
 R:Haanes, E.J.; Heath, D.G.; Cleary, P.P.

J. Bacteriol. 174, 4967-4976, 1992  
A:Title: Architecture of the vir regions of group A streptococci parallels opacity factors  
A:Reference number: A42711; MUID:92332431  
A:Accession: A42711  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 345-415 <HA>  
A:Cross-references: GI:M66806; NID:g153630; PIDN:AAA26687.1; PID:g153631  
A:Experimental source: Strain CS101, OF+  
A:Note: Sequence extracted from NCBI backbone (NCBIN:108942, NCBIPI:108945)  
Superfamily: M5 protein

Query Match	11.1%	Score 99	DB 2	Length 415
Best Local Similarity	26.9%	Pred.	No. 2.2	
Matches 35	Conservative 29	Mismatches 46	Indels 20	Gaps 5

```

0Y 57 LRAVNECEN -VTNHLDOELTBAKGRDVEBAQAT-----CNHVAIAMSLD-----A 104
   ::: : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
184 VKSQDEAKNNAETIEDLKQODASKTEE-IANIQSEATITLENILGSAKHEITJDLQAKLDTATA 242

0Y 105 EKAGQCKAVEELG-----EITLNMHKLQDASAEVRLRREQVLSVAIAKCKYPPS 157
   |||::: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 243 EKAKIESOETITLENILGSAKHEITJDLQAKLDANAEKELQSQAAALEKOLEATKRELAD 302

0Y 158 QDSSSAAPQ 167
   : | |
Db 303 LQAKLAATNQ 312

```

RESULT 10  
MWKW

myosin heavy chain B [similarity] - Caenorhabditis elegans  
M:Contains: myosin ATPase (EC 3.6.1.32)  
C:Species: Caenorhabditis elegans  
C:Date: 13-Jun-1983 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
C:Accession: T20770; T21629; A93558; A93287; A21074; A02992  
R:Kershaw, J.  
submitted to the EMBL Data Library, November 1996  
Reference number: Z19322

A:Accession: T20770  
A:Status: translated from GB/EMBL/DDBT  
A:Molecule type: DNA  
A:Residues: 1-1963 <MIL>  
A:Cross-references: EMBL:Z81499; PIDB:CA804089.1; GSPDB:GN00019; CESP:FLIC3.3  
A:Experimental source: clone FLIC3

A:Residues: 1873-1963 <M13>  
A:Cross-references: GB:V01494; GB:V01049; NID:g6783; P1DN:CAA24738.1; P1D:g6784  
C:Genetics:  
A:Gene: unc-54; CESP:F11C3.3  
A:Map position: 1  
C:Introns: 21/3; 64/3; 111/3; 264/1; 525/3; 951/2; 1747/3; 1819/3; 1894/3  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: actin binding; ATP; coiled coil; hydrolyase; methylated amino acid; muscle  
E:84-775/Domain: myosin motor domain homology <MOT>  
E:174-181/Region: nucleotide-binding motif A (P-loop)  
E:662-684/Region: actin binding #status predicted  
E:766-780/Region: actin binding #status predicted  
F:848-1963/Domain: coiled coil #status predicted <COI>  
F:848-1162/Region: S2  
E:1163-1963/Region: light meromyosin  
E:112/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted  
E:180/Binding site: ATP (Lys) #status predicted  
E:702,712/Active site: Cys #status predicted

Query Match	11.1%	Score 98.5	DB 1	Length 1963
Best Local Similarity	24.7%	Pred No. 12		
Matches 36; Conservative	21;	Mismatches 58;	Indels 31;	Gaps 5

Oy	47	KANEACRBDJGAYMECRNTHLLDLOETPAQGFVDVAOATCNHTYMAALMALSIDEX	106
		: : : : : : : : : : : : : : : :	
Dd	1808	EAEAAALKGCGKVIKYLEGVRELESLDGEGRFPDKANKIIGRADRVRELQFGVDEBK	1867
Oy	107	AQ-----GCKVEELEGITTL-----HKLDASAEVELBAREN	141
		: : : : : : : : : : : : : : : :	
Dd	1868	KNEFRLDDLDIKLQOKTKTORQVEBE-ELANILNQTKQLTTHLEDABERAD--QAEN	1922
Oy	142	QVLISVRT---ADKKTYSSODSSSAA	164
		: : : : : : : : : : : : : : : :	
Dd	1925	SLSKMRSKSRASAVAPGLIOSSASAA	1950

RESULT 11  
A55913  
transcytosis-associated protein p115 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 23-Mar-1995 #sequence\_revision 05-Apr-1995 #text\_change 05-Nov-1999  
C:Accession: A55913  
R:Barroso, M.; Neilson, D.S.; Setul, E.  
Proc. Natl. Acad. Sci. U.S.A. 92, 527-531, 1995  
A:Title: Transcytosis-associated protein (TAP)/p115 is a general fusion factor required for transcytosis  
A:Reference number: A55913; MUID:95132633  
A:Accession: A55913  
A:Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-959 <BAR>  
A:Cross-references: GB:U15589; NID:g558474; PID:M52151.1; PID:g558475  
C:Keywords: membrane fusion; membrane trafficking

Query Match	11.0%	Score 98	DB 2	Length 959
Best Local Similarity	30.1%	Pred. No. 6.1		
Matches 37; Conservative			57; Indels	8; Gaps 3

Oy	60	VMECNVTNHLIJOEITTEKORFOOV-EOAATCRHTYMAIAMSIDAKAGOKKVEELEG	118
		739	IEBLRSHQVLLQSQAENDYITENLRSSQYSGMSEQALATCSPDAE-----QVAMEKQ
Oy	119	EITTLNKLQDASAEVLRRENOVYSVRAD-KKYPSSODSSAAAPOLLIVLLGSLA	177
Db	793	ELSAKSQLCSQSLSTRLOATENELDQAAETLAKSVAVGEESLVTAAATTVDEGRSLA	852

QY	178	LLQ	180
		!!!	
Db	853	LLQ	855

## RESULT 12

```

MMKW1
myosin heavy chain D [similarity] - Caenorhabditis elegans
N:Alternate names: myosin heavy chain I
N:Contains: myosin ATPase (EC 3.6.1.32)
C:Species: Caenorhabditis elegans
C:Date: 28-Feb-1986 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: J21193; J23973; S02772; A02993
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: 219388
A:Accession: J21193
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1938 <MTL>
A:Cross-references: EMBL:J21261; PIDN:CAA95806.1; GSPDB:GNO0019; CESP:R06C7.10
R:Gardner, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: 219825
A:Accession: J23973
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1938 <MT2>
A:Cross-references: EMBL:Z71266; PIDN:CAA95848.1; GSPDB:GNO0019; CESP:R06C7.10
R:Diib, N.J.; Maruyama, I.N.; Krause, M.; Karn, J.
J. Mol. Biol. 205, 603-613, 1989
A:Title: Sequence analysis of the complete Caenorhabditis elegans myosin heavy chain gen
A:Reference number: S02771; MUID:89178677
A:Accession: S02772
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-376, 'V', 378-390, 'V', 392-576, 'L', 578-680, 'I', 682-1938 <DIB>
A:Cross-references: EMBL:X08065; NID:96785; PIDN:CAA30854.1; PID:96786
R:Karn, J.; Brenner, S.; Barnett, L.
Proc. Natl. Acad. Sci. U.S.A. 80, 4253-4257, 1983
A:Title: Protein structural domains in the Caenorhabditis elegans unc-54 myosin heavy ch
A:Reference number: A93958; MUID:83273600
A:Accession: A02993
A:Molecule type: DNA
A:Residues: 24-93, 'E', 95-97, 'R', 99-376, 'V', 378-388, 'GDV', 392-407, 'N', 409-473, 'G', 475-576
C:Genetics:
A:Gene: myo-1; CESP:R06C7.10
A:Map position: 1
A:Introns: 23/3; 114/3; 229/1; 264/1; 320/1; 857/3; 1745/3; 1814/1; 1892/3
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle co
F:87-773/Domain: myosin motor domain homology <MMOT>
F:177-184/Region: nucleotide-binding motif A (P-loop)
F:660-682/Region: actin binding #status predicted
F:764-778/Region: actin binding #status predicted
F:846-1038/Domain: coiled coil #status predicted <COL>
F:846-1160/Region: S2
F:1161-1938/Region: light meromyosin
F:128/Modified site: NP, N6, N6-trimethyllysine (lys) #status predicted
F:183/Binding site: ATP (lys) #status predicted
F:700/710/Active site: Cys #status predicted

Query Match 11.0%; Score 98; DB 1; Length 1938;
Best Local Similarity 26.6%; Pred. No. 13;
Matches 34; Conservative 24; Mismatches 58; Indels 12; Gaps 4;
OY 51 EACRGLRAVME-----CRVTHLLDQ--ELTEAQ-KGPDVAQAATCCHHTVMAALMASLD 103
DB 1324 KAEDELHERQEFHACKNLHELDQCHLEEQINGKDDIOROLSRINSEISOMKARYE 1383
OY 104 AKAGOKKVELEGEITTLNKKLQDASAEVERLRENOVLVRINDKYYRPSODSSA 163
DB 1384 GGGVGSSELEELKKRQKMRVVDLQALSA-----QNKVVISLERAKGKLAEATEDARS 1438
OY 164 AAPOLLIV 171

```

```

DB 1439 VDRHLTVI 1446

RESULT 13
A59293
skeletal myosin heavy chain - domestic rabbit
C:Species: Oryctolagus cuniculus
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000
C:Accession: A59293
R:Maeda, K.; Hostilnova, E.; Roesc. Kleinkauf, A.; Schuster, H.; Gasperik, J.; Wittingh
submitted to GenBank, July 1995
A:Description: Isolation, sequencing of myosin heavy chain cDNA from rabbit skeletal
A:Reference number: A59293
A:Accession: A59293
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1938 <MAE>
A:Cross-references: GB:032574; NID:g940232; PIDN:AA74199.1; PID:g940233
A:Experimental source: strain New Zealand White; cell type skeletal muscle fiber type
C:Genetics:
A:Gene: MHC
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:89-769/Domain: myosin motor domain homology <MMO>

Query Match 11.0%; Score 98; DB 2; Length 1938;
Best Local Similarity 26.1%; Pred. No. 13;
Matches 37; Conservative 20; Mismatches 45; Indels 40; Gaps 5;
OY 47 KANSEACR-----DGLRAVMECRNVTHLLQDELTEAQGFQDYEAQAATCCHHTVMA 99
DB 1366 KANSEVAQWRKTYETDIAQRTLELEAKKKLQRLQDAE---EHVEAVNAK----- 1414
OY 100 ASLDEKAKGQKKVELEGEITTLN-----HKLQDASAEVERLR 138
DB 1415 ASLEKTKRQLQNEVEDLMDIVERTNAACAAALDKORNFEDKILAEWKHKYEETHAELEASQ 1474
OY 139 RENQVLSVRIND-KKYYRPSOD 159
DB 1475 KESRSLSTEVEFKVKNAYESLD 1496

```

```

RESULT 14
S07533
puff II/9A-2 protein precursor - fungus gnat (Sclara coprophila)
C:Species: Sclara coprophila
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 20-Jun-2000
C:Accession: S07533
R:Di Bartolomeis, S.M.; Gerbl, S.A.
J. Mol. Biol. 210, 531-540, 1989
A:Title: Molecular characterization of DNA puff II/9A genes in Sclara coprophila.
A:Reference number: S07532; MUID:90133907
A:Accession: S07533
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-286 <DIB>
A:Cross-references: GB:X51679; NID:g10113; PID:g1405812
C:Genetics:
A:Map position: II/9A
C:Keywords: coiled coil; glycoprotein
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-286/Product: puff II/9A protein #status predicted <MAT>
F:156/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 10.9%; Score 97; DB 2; Length 286;
Best Local Similarity 25.6%; Pred. No. 2; 1;
Matches 31; Conservative 24; Mismatches 36; Indels 30; Gaps 5;
OY 46 IKANSEACDGLRAVMEC-RNVTHL-----LQDELTEAQGFQDYEAQAATCCHHTVMA 98
DB 89 LKREKARQAKKALKKCCQKNTENLKETIEOLKELAEAKALEKCKELADCK----- 142

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: July 18, 2002, 18:38:22 ; Search time 12.96 Seconds  
(Without alignments)  
339.244 Million cell updates/sec

Title: US-09-828-217-1

Perfect score: 889

Sequence: 1 NASTSYDYCRVPMEDGDKRC.....SSAAPOLLIVLGISALLQ 180

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/p1oddata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/p1oddata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/p1oddata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/p1oddata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/p1oddata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/p1oddata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	889	100.0	180	2	US-08-624-650-1
2	99.5	11.2	270	2	US-09-035-095-4
3	99.5	11.2	270	2	US-08-809-494A-2
4	99.5	11.2	270	4	US-09-352-302-2
5	99.5	11.2	273	2	US-08-809-494A-4
6	99.5	11.2	273	4	US-09-352-302-4
7	95	10.7	885	2	US-08-533-306A-4
8	95	10.7	885	2	US-08-742-923A-4
9	93.5	10.5	1886	4	US-08-938-105-3
10	92.5	10.4	292	2	US-08-688-342-4
11	92.5	10.4	292	2	US-09-113-788-4
12	89.5	10.1	316	4	US-09-111-470-4
13	89.5	10.1	1939	4	US-09-310-187A-1
14	88	9.9	288	3	US-08-312-949-4
15	88	9.9	288	3	US-08-446-201-4
16	88	9.9	619	1	US-08-465-746-2
17	88	9.9	619	1	US-08-214-164-2
18	88	9.9	619	2	US-08-467-852A-3
19	88	9.9	619	2	US-08-246-636-3
20	88	9.9	619	2	US-08-247-491A-3
21	88	9.9	619	2	US-08-319-795-2
22	88	9.9	619	2	US-08-468-985-2
23	88	9.9	619	3	US-08-312-949-2
24	88	9.9	648	1	US-08-072-070-2
25	88	9.9	648	1	US-08-469-434-2
26	88	9.9	648	1	US-08-214-222-2
27	88	9.9	648	1	US-08-467-852A-2

28	88	9.9	648	2	US-08-468-718-2	Sequence 2, Appl1
29	88	9.9	648	2	US-08-247-491A-2	Sequence 2, Appl1
30	88	9.9	648	1	US-08-446-201-3	Sequence 2, Appl1
31	88	9.9	695	1	US-08-127-499A-23	Sequence 23, Appl1
32	88	9.9	695	1	US-08-482-847-23	Sequence 23, Appl1
33	87.5	9.8	289	1	US-08-072-070-4	Sequence 4, Appl1
34	87.5	9.8	289	1	US-08-469-434-4	Sequence 4, Appl1
35	87.5	9.8	289	1	US-08-214-222-4	Sequence 4, Appl1
36	87.5	9.8	289	2	US-08-467-852A-5	Sequence 5, Appl1
37	87.5	9.8	289	2	US-08-468-718-4	Sequence 4, Appl1
38	87.5	9.8	289	2	US-08-247-491A-5	Sequence 5, Appl1
39	87.5	9.8	477	1	US-08-402-217A-3	Sequence 3, Appl1
40	87.5	9.8	477	1	US-08-700-178-3	Sequence 3, Appl1
41	87.5	9.8	477	3	US-08-995-654-3	Sequence 3, Appl1
42	87.5	9.8	667	4	US-09-071-709-9	Sequence 9, Appl1
43	85.5	9.6	459	4	US-09-071-709-1	Sequence 1, Appl1
44	85	9.6	344	6	5210183-2	Patent No. 5210183
45	85	9.6	683	6	5210183-3	Patent No. 5210183

## ALIGNMENTS

RESULT 1  
US-08-624-650-1  
Sequence 1, Application US/08624650  
Patent No. 5914252  
GENERAL INFORMATION:  
APPLICANT: HIRANO, TOSHIO  
APPLICANT: KAISHO, TSUNEYASU  
TITLE OF INVENTION: MEMBRANE PROTEIN POLYPEPTIDE HAVING  
TITLE OF INVENTION: PRE-B CELL GROWTH-SUPPORTING ABILITY AND A GENE THEREOF  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
ADDRESSEE: P.C.  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/624,650  
FILING DATE: 22-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP94/01732  
FILING DATE: 14-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-281622  
FILING DATE: 15-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 7625-001-0 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 180 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-624-650-1

Query Match 100.0%; Score 889; DB 2; Length 180;  
Best Local Similarity 100.0%; Pred. No. 4.3e-88;  
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASTSDYCRVPEMDGDKCKLLGIGIIVLITVLIGVPLIIFTKANSEACRDLRAV 60  
DB 1 MASTSDYCRVPEMDGDKCKLLGIGIIVLITVLIGVPLIIFTKANSEACRDLRAV 60

QY 61 MECRNVTLLQOELTEAOKGFQDVEAQAATCNHTVAMALASDAEKAQCKKVELEGET 120  
DB 61 MECRNVTLLQOELTEAOKGFQDVEAQAATCNHTVAMALASDAEKAQCKKVELEGET 120

QY 121 TTLNHLKODASAEVRLRENOVLSVRIADKKYPPSSQDSSAAPOLLIVLIGSALIQ 180  
DB 121 TTLNHLKODASAEVRLRENOVLSVRIADKKYPPSSQDSSAAPOLLIVLIGSALIQ 180

RESULT 2  
US-09-055-095-4  
Sequence 4, Application US/09055095  
Patent No. 5945308  
GENERAL INFORMATION:  
APPLICANT: Tan9, Y. Tom  
APPLICANT: Cortley, Chandra  
APPLICANT: Cortley, Neil C.  
APPLICANT: Sather, Susan  
TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/055,095  
FILING DATE: Filed Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0500 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 270 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1902982  
US-09-055-095-4

Query Match 11.2%; Score 99.5; DB 2; Length 270;  
Best Local Similarity 25.9%; Pred. No. 0.0068;  
Matches 36; Conservative 25; Mismatches 41; Indels 37; Gaps 5;

QY 23 LIGIGIVLITVLIGVPLIIFTKANSEACRDLRAVMECRNVTHLLOELTEAOKGFQ 82  
DB 40 VLGIGIVLITVLIGVPLIIFTKANSEACRDLRAVMECRNVTHLLOELTEAOKGFQ 82

QY 83 DVEAQAATCNHTVAMALASDAEKAQCKKVELEGETITTLNHLKODASAEVRLRENO 142  
DB 81 DVEAQAATCNHTVAMALASDAEKAQCKKVELEGETITTLNHLKODASAEVRLRENO 142

QY 143 VLSVRIADKKY--PSSOD 159  
DB 125 NLOEVLKEAANYSGPCPD 143

RESULT 3  
US-08-809-494A-2  
Sequence 2, Application US/08809494A  
Patent No. 5962260  
GENERAL INFORMATION:  
APPLICANT: Masaki, Tatsuya  
APPLICANT: Sawamura, Tomoo  
TITLE OF INVENTION: Modified Low-Density Lipoprotein  
TITLE OF INVENTION: Receptor  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McAlay Fisher Nissen Goldberg & Kiel  
STREET: 261 Madison Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10016-2391  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/809,494A  
FILING DATE: 24-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-321705  
FILING DATE: 30-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-214206  
FILING DATE: 31-JUL-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldberg, Jules E.  
REGISTRATION NUMBER: 24408  
REFERENCE/DOCKET NUMBER: JG-YY-4363PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 986-4090  
TELEFAX: 212 818-9479  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 270 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-809-494A-2

Query Match 11.2%; Score 99.5; DB 2; Length 270;  
Best Local Similarity 25.9%; Pred. No. 0.0068;  
Matches 36; Conservative 25; Mismatches 41; Indels 37; Gaps 5;

QY 23 LIGIGIVLITVLIGVPLIIFTKANSEACRDLRAVMECRNVTHLLOELTEAOKGFQ 82  
DB 40 VLGIGIVLITVLIGVPLIIFTKANSEACRDLRAVMECRNVTHLLOELTEAOKGFQ 82

QY 83 DVEAQAATCNHTVAMALASDAEKAQCKKVELEGETITTLNHLKODASAEVRLRENO 142  
DB 81 DVEAQAATCNHTVAMALASDAEKAQCKKVELEGETITTLNHLKODASAEVRLRENO 142

QY 143 VLSVRIADKKY--PSSOD 159  
DB 125 NLOEVLKEAANYSGPCPD 143

```

      APPLICANT: Masaki, Tomoo
      TITLE OF INVENTION: Modified Low-Density Lipoprotein
      NUMBER OF SEQUENCES: 8
      CORRESPONDENCE ADDRESS:
        ADDRESSEE: McAulay Fisher Nissen Goldberg & Kiel
          STREET: 261 Madison Avenue
            CITY : New York
              STATE: NY
                COUNTRY: USA
                  ZIP: 10016-2391
                    COMPUTER READABLE FORM:
                      MEDIUM TYPE: Floppy disk
                        COMPUTER: IBM PC compatible
                          OPERATING SYSTEM: PC-DOS/MS-DOS
                            SOFTWARE: PatentIn Release #1.0, Version #1.30
                              CURRENT APPLICATION DATA:
                                APPLICATION NUMBER: US/08/809,494A
                                  FILING DATE: 24-MAR-1997
                                    CLASSIFICATION: 435
                                      PRIOR APPLICATION DATA:
                                        APPLICATION NUMBER: JP 6-321705
                                          FILING DATE: 30-NOV-1994
                                            PRIOR APPLICATION DATA:
                                              APPLICATION NUMBER: JP 7-214206
                                                FILING DATE: 31-JUL-1995
                                                  ATTORNEY/AGENT INFORMATION:
                                                    NAME: Goldberg, Jules E
                                                      REGISTRATION NUMBER: 24408
                                                        REFERENCE/DOCKET NUMBER: JG-YJ-4363PCT
                                                          TELECOMMUNICATION INFORMATION:
                                                            TELEPHONE: 212 986-4090
                                                              TELEFAX: 212 818-9479
                                                                INFORMATION FOR SEQ ID NO: 4:
                                                                  SEQUENCE CHARACTERISTICS:
                                                                    LENGTH: 273 amino acids
                                                                      TYPE: amino acid
                                                                        TOPOLOGY: linear
                                                                          MOLECULE TYPE: protein
                                                                            US-08-809-494A-4

Query Match               11.2%; Score 99.5; DB 2; Length 273;
Best Local Similarity     25.9%; Pred. No. 0.0069;
Matches    36; Conservative   25; Mismatches    41; Indels    37; Gaps       5.
QY      23 LLGIILVLIIIVLIGVPDITFIPTIKANSPACRDLRAVMESGNVTHTLDQEALTEAQKGR 82
         ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      43 VLCLGLTAVTLIQLDSVDLIKQQ-----ANTH--OEDILEGO---- 83
QY      83 DVEAGATCNMTHVALMLASLDAEKAQGOKYELEGETITLNHHKLQODASAEEVERLLRENO 142
         ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      84 -----ILNQRRSEKS-AQESCKELNEMITTLANHKIDSKSKIMELHRNL 127
QY      143 VLSVRADKKYY--PSSD 159
         |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      128 NLQEVLRKANYSGPCPD 146
RESULT           6
US-09-352-302-4
Sequence 4, Application US/09352302
Patent No. 6197937
GENERAL INFORMATION:
APPLICANT: Sawamura, Tatsuya
APPLICANT: Masaki, Tomoo
TITLE OF INVENTION: Modified Low-Density Lipoprotein
TITLE OF INVENTION: Receptor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McAulay Fisher Nissen Goldberg & Kiel
STREET: 261 Madison Avenue
CITY: New York
```

QY	53	CRDGLRAMECRNRYTHLLQOELTEKGFQDYEAQATCNTHVMALMSLD-----	103
Db	215	CSDGRRRAELNDKYKHLQNEEVSATGMLNEEGGAIKIAKAVASISLSDQDTQELDDE	274
QY	104	-AEKAGQKVEELTEGELITTLNHLKLDQASAEYERLRRENQVLSVRIADKK	152
Db	275	TROKLNASTKRLQTEEERNSSLDQDLEDEEAKQKNERHISTINIDLSDK	324
		:	

Sequence 4, Application US/08742923A  
Patent No. 5869611  
GENERAL INFORMATION:  
APPLICANT: Liu, Pu  
APPLICANT: Collins, Francis S.  
APPLICANT: Siciliano, Michael J.  
APPLICANT: Claxton, David  
TITLE OF INVENTION: Markers for Detection of Chromosome 16  
REARRANGEMENTS  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.  
STREET: P.O. Box 828  
CITY: Bloomfield Hills

```

STATE: MI
COUNTRY: USA
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,923A
FILING DATE: No. 5869611ember 1, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Deann F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869DVC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 885 amino acids

```

QY 161 SSAAPQLLIVLGL 175

APPLICANT: Au-Young, Janice

APPLICANT: Au-Young, Janice

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

? Patent No. 6042838
? GENERAL INFORMATION:
? APPLICANT: BILLES, David E.
? APPLICANT: WU, Hong-Yin
? TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS FOR MUCOSAL ADMINISTRATION OF
? FILE REFERENCE: 454312-2018
? CURRENT APPLICATION NUMBER: US/08/446,201B
? EARLIER FILING DATE: 1995-05-19
? EARLIER APPLICATION NUMBER: 08/312,949
? EARLIER FILING DATE: 1994-09-30
? EARLIER APPLICATION NUMBER: 08/246,636
? EARLIER FILING DATE: 1994-05-20
? EARLIER APPLICATION NUMBER: 08/048,896
? EARLIER FILING DATE: 1993-04-20
? EARLIER APPLICATION NUMBER: 07/835,698
? EARLIER FILING DATE: 1992-02-12
? EARLIER APPLICATION NUMBER: 07/656,773
? EARLIER FILING DATE: 1991-02-15
? NUMBER OF SEQ ID NOS: 4
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 4
? LENGTH: 288
? TYPE: PR1
? ORGANISM: Streptococcus pneumoniae
? SS-08-446-201-4

```

```

Query Match          9.9%; Score 88; DB 3; Length 288;
Best Local Similarity 30.9%; Pred. NO. 0.13;
Matches 34; Conservative 15; Mismatches 29; Indels 32; Gaps 5

QY      60  VMECHNVTHLLOEELTE-----AOKGFODVBAQATCNHTYMALASDAEKAQOQ 110
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 178  IAELENOVHRLRLOELKEIDSESESDYEAKEGFR-----APLOSKIDAKKAKLS 224

QY      111 K-----KVEELEGEITTLNHRKLODA--SAEYERLRRENOVLSVRIADKK 152
      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 225  KLEELSDKIDELDAETIAKLDELQKPAEERNNNVEDYFKEG--LEKTIAMKK 272

Search completed: July 18, 2002, 18:42:58
Job time: 276 sec

```

THIS PAGE BLANK (USPTO)